THE UNIVERSITY
of EDINBURGH

Genome editing using site-specific nucleases: Targeting highly expressed genomic regions for robust transgene expression and genetic analysis

Peter Andrew Tennant

Thesis submitted for the degree of Doctor of Philosophy

The University of Edinburgh

2015

Declaration

I declare,

(a) that this thesis has been composed by myself, and
(b) that this work is my own, unless otherwise stated, and
(c) that this work has not been submitted for any other degree or professional qualification.

Peter Andrew Tennant

May 2016

Acknowledgements

It feels like only yesterday that I embarked upon my PhD, but alas, four years have passed and it has drawn to an end. I have thoroughly enjoyed my time, but none of this would have been possible without the support and encouragement of a good number of people, whom I shall endeavour to thank here.

Firstly, I would like to thank my supervisor Chris Boyd, both for agreeing to take me on as a PhD student and for providing support and advice throughout. I want to thank you for challenging me intellectually and allowing me to find my own path. I would not have been able to complete this project without you, so thank you. I would also like to thank Andrew Wood, Nick Gilbert, and Sally Cross for their help and advice, and for always finding the time to help me when I asked. Special thanks go to Cathy Abbott for offering me endless help and support, for being an excellent thesis committee chair, and for allowing me to gate-crash your lab meetings, in order to gain a greater understanding of the eEF1A field.

I must also thank all the members of the CF group past and present, James, Javi, June, Barbie, Ann, Heather, and Rob, for their support, and for making me feel welcome and part of the team. A particular thank you goes to Laura H. for looking after me in the lab, generally putting up with my endless questions and, most importantly, for being a good friend.

The CGEM (and IGMM as a whole) has some excellent students, I must thank my office friends in particular, Laura M., Faith, and Daniel, for all the useful and random conversations and for keeping me entertained for the past four years. I must also thank Faith for teaching me the dark art of western blotting and Daniel for all his help with qPCR. I would also like to thank all the other students, especially Nidhi, Kate, and Katie, but many more, for helping to make my time here an enjoyable one.

There are so many other people in the CGEM who have helped make my experience a great one, too many to name, but I would like to thank Susan, Helen, Rosemary, Joyce, and Jennifer in particular, for being so welcoming and for always giving me help and advice.

A special thank you goes to Niamh, for your wisdom, but also for being a great friend and an excellent flatmate. This last year would not have been the same without you, long may the craziness continue. I would also like to thank my good friend Neil, for your words of encouragement, for always being there when I needed to moan, for your proofreading skills, for providing just enough of a distraction with the DDD and for pints in even numbers, thank you. I must also thank my friends Tom and Kirstie, you were always there for me when I needed you and offered me lots of support along the way.

Finally, I want to thank my parents, for always believing in me and supporting me, for always being there when I needed a chat or a cheer up and generally for being truly wonderful people, thank you.

Abstract

Integration and expression of exogenous genetic material – in particular, transgenes – into the genomes of model organisms, cell lines or patients is widely used for the creation of genetically modified experimental systems and gene therapy. However, loss of transgene expression due to silencing is still a major hurdle which remains to be overcome. Judicious selection of integration loci can help alleviate the risk of silencing; in recent years the ability to efficiently and specifically target transgene integration has been improved by the advent of site-specific nucleases (SSNs). SSNs can be used to generate double strand breaks (DSBs) in a targeted manner, which increases the efficiency of homologous recombination (HR) mediated transgene integration into predetermined loci. In this work I investigate four human genomic loci for their potential to act as transgene integration sites which will support robust long term expression: the adeno-associated virus (AAV) integration site 1 (AAVS1); the human homologue of the mouse *Rosa26* locus (hROSA26); the inosine monophosphate dehydrogenase 2 (*IMPDH2*) gene and the eukaryotic translation elongation factor 1 alpha 1 (*EEF1A1*) gene. I also investigate the potential of creating a novel drug-selectable transgene integration system at the *IMPDH2* locus to allow for rapid and specific selection of correctly inserted transgenes.

In addition to their ability to drive targeted transgene integration, SSNs can be harnessed to specifically disrupt gene function through indel formation following erroneous repair of the induced DSB. Using this strategy, I aimed to answer some important biological questions about eukaryotic translation elongation factor 1 alpha (eEF1A); eEF1A is responsible for providing aminoacylated tRNAs to the ribosome during the elongation phase of protein synthesis. Humans and other vertebrates express two isoforms, eEF1A1 and eEF1A2 (encoded by *EEF1A1* and *EEF1A2* respectively). During development eEF1A1 is replaced by eEF1A2 in some tissues. The reasons for this remain elusive, but one explanation may lie in the moonlighting functions of eEF1A1, which may not be shared by eEF1A2. Additionally, eEF1A2 can act as an oncogene, while there is no evidence that eEF1A1 is overexpressed in tumours. To begin to untangle these issues I targeted *EEF1A1* using SSNs with the aim of making a cell line expressing only the eEF1A2 isoform. This work suggests

that eEF1A1 may be essential even in the presence of eEF1A2, though further studies will be required to confirm this.

Lay Summary

The ability to insert genes into the genome is an essential technique in cell biology, both for studying gene function, making disease models and ultimately for gene therapy, whereby therapeutic genes are delivered to patients to replace the function of faulty genes. However, this technique can be hampered when these genes become switched off by a process known as silencing. In order to avoid this, a targeted approach can be used, whereby safe sites in the genome can be chosen and used to specifically integrate the gene. The ability to do this has been greatly improved by the use of genetically engineered enzymes called site-specific nucleases. These work like molecular scissors which can be made to cut chromosomes at specified positions within the genome and vastly increase the efficiency of gene insertion. In this work, I have been using this technology to investigate four chromosomal sites in the human genome for their potential to support the insertion of genes, without becoming switched off or silenced.

In addition to targeting gene insertion, site-specific nucleases can be used to make gene deletions. I have been using this property to try and delete the *EEF1A1* gene in cells. This gene encodes a protein called eEF1A1, which plays a vital role in the cell's protein production machinery. All cells in the body have eEF1A1 except a few specialised cells, such as those in muscles, the heart and some nerves, where a highly similar protein – eEF1A2 – performs this role instead. The reasons why we have two almost identical proteins, performing the same role remain unknown, but it may be due to differences in their other cellular functions. It is known that eEF1A1 performs many other roles in the cell, but it is unknown whether eEF1A2 shares these. In addition, eEF1A2 is known to play a role in cancer, while this is not the case for eEF1A1. I sought to delete eEF1A1 in order to begin to unravel the reasons for this, and to allow me to look at the functions of eEF1A2 in more detail. This work seems to indicate that eEF1A1 may be essential for cell survival even in the presence of eEF1A2, though further work will be required to confirm this.

Contents

List of Figures

List of Tables

List of Abbreviations

°C	Degrees centigrade
µ	*Prefix* micro
6-TG	6-Thioguanine
A1AT	α1-antitrypsin
aa-tRNA	Aminoacyl-tRNA
AAD	Acidic activation domain
AAV	Adeno-associated virus
AAVS1	Adeno-associated virus integration site 1
amp	Ampicillin
AMPS	Ammonium persulfate
ANOVA	Analysis of variance
ATCC	American type culture collection
ATP	Adenosine triphosphate
attB	Bacterial attachment
attP	Phage attachment
AZT	Azidothymidine
BAC	Bacterial artificial chromosome
BCA	Bicinchoninic acid
BET	Bromodomain and extra terminal
bp	Base pair
BrdU	Bromodeoxyuridine
BSA	Bovine serum albumin
Cas9	CRISPR associated 9
Cas9n	Cas9 nickase
CBX3	Chromobox protein homologue 3
CCR5	Chemokine (CC motif) receptor 5
cDNA	Complementary DNA
CEBPA	CCAAT/enhancer binding protein (C/EBP), alpha
CEBPB	CCAAT/enhancer binding protein (C/EBP), beta
CF	Cystic fibrosis
CFTR	Cystic fibrosis transmembrane conductance regulator
CGI	CpG Island
CHD1	Chromodomain helicase DNA-binding protein 1
CHO	Chinese hamster ovary
cHS4	Chicken hypersensitive site 4
CMV	Cytomegalovirus
CRISPR	Clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
Ct	Cycle threshold
CTCF	CCCTC-binding factor
del	Deletion
dGDP	Deoxyguanosine diphosphate
dGTP	Deoxyguanosine triphosphate
dH_2O	Distilled water
DMEM	Dulbecco's modified eagle medium
DMSO	Dimethyl sulfoxide

DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
dNTP	Deoxynucleotide triphosphate
ds	Double stranded
DSB	Double strand break
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EB	Elution buffer
E-box	Enhancer box
EDTA	Ethylenediaminetetraacetic acid
eEF	Eukaryotic translation elongation factor
eEF1A	Eukaryotic translation elongation factor 1 alpha
EGFP	Enhanced green fluorescent protein
EGR	Early growth response
ELISA	Enzyme linked immunosorbent assay
ES cells	Embryonic stem cells
Exp5	Exportin 5
FACS	Fluorescence activated cell sorting
FCS	Foetal calf serum
FLASH	Fast ligation-based automatable solid-phase high-throughput
FOXH1	Forkhead box H1
FRT	Flippase recognition target
g	Gram
gDNA	Genomic DNA
GDP	Guanosine diphosphate
GMP	Guanosine monophosphate
GMPS	GMP synthase
gRNA	Guide RNA
GTP	Guanosine triphosphate
HAT	Histone acetyl transferase
HAT	Hypoxanthine-aminopterin-thymidine
HBB	β-goblin gene
HBD	δ-globin gene
HBx	Hepatitis B virus protein X
HCC	Hepatocellular carcinoma
HDAC	Histone deacetylase
HDR	Homology directed repair
HIRI	Homologous illegitimate random integration
HIV	Human immunodeficiency virus
HNRPA2B1	Human heterogeneous nuclear riboprotein A2 B1
HP1	Heterochromatin protein 1
HPRT	Hypoxanthine-guanine phosphoribosyltransferase
HR	Homologous recombination
IGF2	Insulin-like growth factor 2
IL-2	Interleukin 2
IL-2Rγ	Interleukin 2 receptor gamma chain
IMP	Inosine monophosphate
IMPDH	Inosine monophosphate dehydrogenase

Indel	Insertion or deletion
ins	Insertion
iPSC	Induced pluripotent stem cell
IR	Inverted repeat
Kb	Kilobase
KCl	Potassium chloride
KDa	Kilodalton
KH_2PO_4	Monopotassium phosphate
KLD	Kinase/ligase/DpnI
L	Litre
LB	Luria broth
LMO2	LIM-domain only 2
LTR	Long terminal repeat
M	Molar
m	*Prefix* milli
m5C	5-methylcytosine
MeCP2	Methyl CpG binding protein 2
$MgCl_2$	Magnesium chloride
MLV	Murine leukaemia virus
MMF	Mycophenolate mofetil
MMR	Mismatch repair
MPA	Mycophenolic acid
MTA3	Metastasis associated 1 family, member 3
MYST	MOZ, Ybf2/Sas3, Sas2 and Tip60
n	*Prefix* nano
Na_2HPO_4	Sodium phosphate dibasic
NaCl	Sodium chloride
NAD^+	Nicotinamide adenine dinucleotide
NFIC	Nuclear factor I/C (CCAAT-binding transcription factor)
NF-water	Nuclease free water
NGS	Next generation sequencing
NHEJ	Non-homologous end joining
NK	Natural killer
NK	Nucleotide kinase
NLS	Nuclear localisation signal
ns	Not significant
nt	Nucleotide
NTC	No template control
ODN	Oligodeoxynucleotide
PAC	P1-derived artificial chromosome
PAGE	Polyacrylamide gel electrophoresis
PAM	Protospacer adjacent motif
PB	PiggyBac
PBS	Phosphate buffered saline
PBS-T	Phosphate buffered saline/Tween-20
PCR	Polymerase chain reaction
PE	Position effect
PEV	Position effect variegation

PI3K	Phosphoinositide 3-kinase
PNK	Polynucleotide kinase
PNP	Purine nucleoside phosphorylase
POLR2A	polymerase (RNA) II (DNA-directed) polypeptide A
PPP1R12C	Protein phosphatase 1, regulatory (inhibitor) subunit 12C
pre-crRNA	Pre-CRISPR RNA
PTM	Post-translational modification
qPCR	Quantitative PCR
RDEB	Recessive dystrophic epidermolysis bullosa
RIGS	Repeat-induced gene silencing
RIN	RNA integrity number
RMCE	Recombinase-mediated cassette exchange
RNA	Ribonucleic acid
rpm	Revolutions per minute
RVD	Repeat variable di-residue
S. pyogenes	*Streptococcus pyogenes*
SB	Sleeping Beauty
SCID-X1	X-linked severe combined immunodeficiency
SDS	Sodium dodecyl sulphate
S.E.M.	Standard error of the mean
sgRNA	Single guide RNA
SIN	Self-inactivating
siRNA	Small interfering RNA
SNP	Single nucleotide polymorphism
SOX10	SRY (sex determining region Y)-box 10
SOX9	SRY (sex determining region Y)-box 9
SRY	Sex determining region Y
SSB	Single strand break
ss	Single stranded
SSN	Site-specific nuclease
STAT3	Signal transducer and activator of transcription 3
STAT5A	Signal transducer and activator of transcription 5A
sub	Substitution
T. foetus	*Tritrichomonas foetus*
TALE	Transcription activator-like effector
TALEN	Transcription activator-like effector nuclease
TBE	Tris/Borate/EDTA
TEMED	Tetramethylethylenediamine
TF	Transcription factor
Tf	Tritrichomonas foetus
TGS	Tris/Glycine/SDS
THAP1	THAP domain containing, apoptosis associated protein 1
Tm°	Annealing temperature
tracrRNA	Trans-activating crRNA
tRNA	Transfer RNA
TRD	Transcriptional repressor domain
UC	Untransfected control
UCOE	Ubiquitous chromatin opening element

UCSC	University of California, Santa Cruz
UKCFGTC	United Kingdom CF gene therapy consortium
UTR	Untranslated region
V	Volts
VSV-G	Vesicular stomatitis virus G protein
WT	Wild type
XMP	Xanthosine monophosphate dehydrogenase
XPRT	Xanthine phosphoribosyl transferase
YAC	Yeast artificial chromosome
YYI	YY1 Transcription Factor
ZFN	Zinc finger nuclease
ZnF	Zinc finger

Chapter 1 Introduction

Recent advances have dramatically enhanced our ability to specifically alter DNA sequences *in vitro* and *in vivo*, a process known as gene editing (Carroll 2014). However, the ability to alter the genome by the insertion of exogenous DNA, particularly transgenes, is a vital technique in cell biology. It has allowed investigators to explore fundamental biological questions and has been instrumental in the production of cellular disease models and transgenic animals. It has also been useful for the development of recombinant proteins both for research and pharmaceutical applications and will have far-reaching implications in the field of gene therapy.

This work will focus on the use of gene editing technologies for the stable integration and long term expression of transgenes in human cell lines. This will provide a proof of principle for the production of recombinant lentiviral producer cell lines, for use in gene therapy applications.

1.1 Gene Silencing

A challenge facing all transgenic studies is the loss of transgene expression due to silencing. Under normal circumstances, gene silencing is necessary for cellular function. Gene expression is not determined by primary DNA sequence alone. In addition, it is profoundly influenced by epigenetic factors, notably proteins that interact with DNA and each other with short- and long-term effects on transcription. Methylation of cytosine residues at CpG dinucleotides also plays an important role.

However, when exogenous DNA sequences, for example, transgenes, become incorporated into the genome, they can be prone to the influence of these factors leading to loss of expression. The exact mechanisms of transgene silencing are not fully understood, but certain processes are known to be involved. I will begin by discussing the structure of chromatin to set the scene for a survey of transgene silencing mechanisms.

1.1.1 Chromatin Structure

The nuclear DNA of eukaryotes is organised into a higher order structure termed chromatin. Chromatin in its lowest order structure is composed of repeating units – nucleosomes – made up of 146-bp of DNA wrapped around an octamer of histone

proteins, two each of H2A, H2B, H3 and H4 (Luger et al., 1997; Figure 1.1). In addition, a fifth histone, H1, interacts with the DNA where it enters and exits the nucleosome core (Kamakaka and Biggins, 2005; Figure 1.1). These histones – in particular their tail regions – are subject to post-translational modifications (PTMs) which can alter the overall structure of the chromatin.

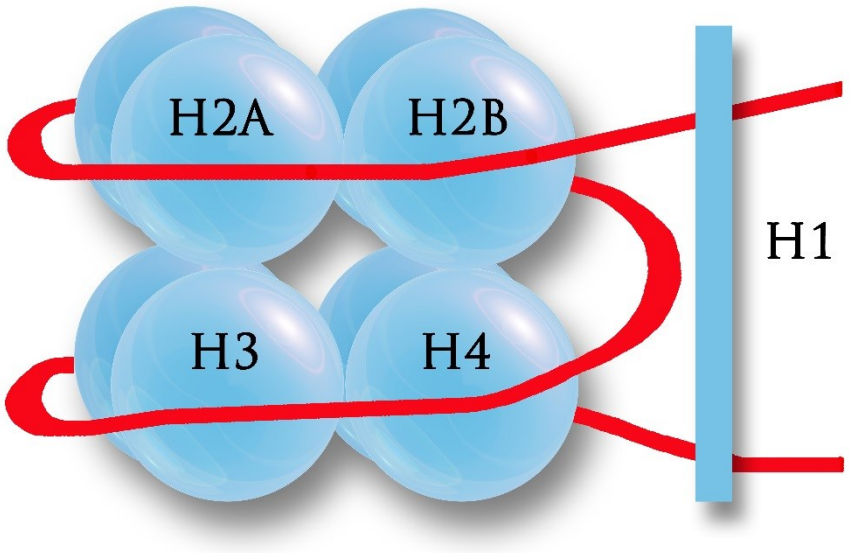


Figure 1.1 Structure of a nucleosome. Each nucleosome is composed of 146-bp of DNA (red) wrapped around an octamer of histone proteins (blue circles). The octamer is composed of two molecules each of H2A, H2B, H3, and H4. Where the DNA enters and exists the nucleosome, it interacts with another histone protein, histone H1 (blue rectangle). Nucleosomes represent the lowest order of chromatin structure, forming the 10 nm fibre, often referred to as 'beads on a string'. This is further compacted to form the 30 nm fibre, termed the solenoid structure. This undergoes several more layers of organisation to form the chromosome.

1.1.2 Histone Modifications

Histone proteins are each composed of a globular domain and an N-terminal tail. The tails are highly basic and protrude from the nucleosomes where they can interact with neighbouring nucleosomes. These domains, and to some extent the globular core domains, are subject to myriad PTMs. These can affect chromatin structure in two ways: 1) they can directly alter the overall charge and thus affect both histone-DNA and inter-nucleosome interactions, and 2) they can recruit enzyme complexes which can either directly remodel positioning of the nucleosomes or regulate histone PTMs (Bannister & Kouzarides 2011). These modifications commonly take the form of acetylation and methylation, but other PTMs can also occur, for example phosphorylation and ubiquitination.

Histone acetylation occurs on lysine residues within the histones, resulting in a neutralisation of their positive charge and a weakening of their interaction with the negatively charged DNA (Bannister & Kouzarides 2011). Within the tail region, acetylation commonly occurs at lysines 9 and 14 of histone H3 (H3K9ac and H3K14ac).

Methylation occurs at both lysine and arginine residues, predominantly in the histone tail regions. Unlike acetylation, methylation does not affect the charge of the histone. Instead, methylated residues can recruit proteins (or protein complexes) which can either stabilise chromatin structure or bring about further changes. For example tri-methylation of lysine 9 on histone H3 (H3K9me3), which recruits heterochromatin protein 1 (HP1) (Lachner et al. 2001).

The combination of these multiple histone modifications in *cis* or *trans* constitute the 'histone code' which can be interpreted by proteins or protein complexes to bring about further changes to the chromatin structure (Strahl & Allis 2000). Proteins are able to recognise these modifications via specific domains, for example the bromodomain is believed to be involved in the recognition and binding of acetylated lysine residues (Dhalluin et al. 1999).

1.1.3 Chromatin Organisation

Chromatin can be divided into two groups based on its structure: euchromatin (open and transcriptionally active) and heterochromatin (closed and transcriptionally repressed). Each type of chromatin is characterised by the presence of different histone modifications. H3K9ac, H3K14ac and tri-methylation of lysine 4 of histone H3 (H3K4me3) are markers of euchromatin. Acetylation neutralizes the positive charge of the lysine residues leading to weakened interactions thus a relaxation of the chromatin structure, which is more amenable to transcription (Bannister & Kouzarides 2011).

Heterochromatin is hypoacetylated and is enriched for DNA methylation at CpG sites (discussed in 1.1.4 below; Barski et al., 2007; Decottignies, 2014; Saksouk et al., 2015). Heterochromatin can be subdivided into constitutive heterochromatin, found in centromeres and telomeres, or facultative heterochromatin often containing genes, which become silenced during development (Saksouk et al. 2015); each type contains a specific pattern of histone PTMs. Constitutive heterochromatin is characterised by the presence of H3K9me3 and di- and tri-methylation of lysine 20 in histone H4 (H4K20me2/3), which is the hallmark of pericentromeric heterochromatin (Saksouk et al. 2015). On the other hand, facultative heterochromatin is enriched for tri-methylation of histone H3 lysine 27 (H3K27me3) and ubiquitination of lysine 199 of histone H2A (H2AK199ub; Saksouk et al., 2015).

Heterochromatin is associated with many protein complexes, which help maintain and stabilise its structure, but can also initiate the spreading of heterochromatin into neighbouring regions. At centromeres it has been shown that H3K9me3 is a crucial mediator of the heterochromatic state; proteins containing a chromodomain, such as HP1, are able to bind to H3K9me3 (Dillon & Festenstein 2002). The enzyme SUV39H is responsible for the majority of H3K9me3 at pericentromeric regions and is also involved in DNA methylation via recruitment of DNA methyltransferases (DNMTs; Muramatsu et al., 2013). In addition, H3K9me3 is required for the recruitment of SUV420H which leads to the di- and tri-methylation of H4K20 (H4K20me2/3), another hallmark of pericentromeric heterochromatin, which can also interact with HP1 (Saksouk et al. 2015).

1.1.4 DNA Methylation

In addition to alterations in the structure of chromatin discussed above, DNA modification can also affect transcriptional activation or repression. In many animals, plants, and fungi, methylation of DNA only occurs at cytosine residues which are often adjacent to guanosine residues in sites termed CpG sites (Suzuki & Bird 2008). This is regulated by DNMTs, which convert cytosine bases (found in CpG sites) into 5-methylcytosines (m5C). Three DNMTs have to date been described: DNMT1, DNMT3A and DNMT3B (Saksouk et al. 2015). Generally, the human genome is hypermethylated except at certain regions termed CpG islands (CGIs), which are often found in the promoter regions of genes (Ioshikhes and Zhang 2000). Methylation at these CGIs or in promoters is associated with gene silencing (Suzuki & Bird 2008).

1.2 Transgene Silencing Mechanisms

1.2.1 Position Effects

Silencing of transgenes can occur as a direct consequence of their integration site, via phenomena known as position effects (PEs); integration into, or close to heterochromatic regions can result in transgene silencing due to PEs. PEs can be stable, for example, when integration occurs at subcentromeric regions, or variable between cells, known as position effect variegation (PEV).

When transgenes integrate into heterochromatic regions in the vicinity of telomeres or centromeres they can be subject to silencing due to the spreading of the heterochromatin found in these regions. These regions of constitutive heterochromatin are marked by the presence of H3K9me3 which is bound by HP1 (Dillon & Festenstein 2002). HP1 can recruit histone methyltransferases such as SUV39H, capable of methylating H3K9 (Muramatsu et al. 2013), which in turn can be recognised by HP1 and thus heterochromatin can propagate and silence the transgene expression (Decottignies 2014).

PEV is caused by incomplete transgene silencing; transgenes (or, in fact, genes generally) become silenced in some, but not all, cells leading to a variegated phenotype. It was first identified in Drosophila, which showed a variegated phenotype following x-ray irradiation (Muller 1930). This resulted from X-ray

induced chromosomal rearrangements, which placed the genes involved adjacent to heterochromatic regions, leading to silencing.

Variegated expression can also result from the integration of tandem arrays of the transgene, which often leads to extinction of transcription, a phenomenon referred to as repeat-induced gene silencing (RIGS). In their 1998 study, Garrick and colleagues demonstrated that transgenes inserted in a concatemeric array showed reduced expression resulting in a variegated phenotype which was reversed when transgene number was reduced. They demonstrated that silencing resulted from hypermethylation of the transgenes and adoption of a repressive chromatin local environment. However, it seems that the link between RIGS and PEV is a complex one and reduction of transgene copy number does not always alleviate PEV. In their study, Williams et al. (2008) demonstrated that the effect of reducing transgene copy number on silencing is dependent on integration site. In addition, for some integration sites, multiple transgene copies may be advantageous. It is likely that a complex interplay between the local environment and the transgene ultimately determines whether it will be silenced.

1.2.2 DNA Methylation of Transgenes

Transgenes that become silenced are associated with increased DNA methylation (Li et al. 2001; Pikaart et al. 1998); studies have shown that inhibition of DNMTs results in reversal of gene extinction (He et al. 2005; Pikaart et al. 1998). However, it has been shown that DNA methylation, whilst important to maintaining the silenced phenotype, may not be a causative event (Mutskov & Felsenfeld 2004). Methylated DNA can be bound by proteins such as methyl CpG binding protein 2 (MeCP2), which contains a transcriptional repressor domain (TRD), but can also recruit protein complexes containing histone deacetylases (HDACs), capable of deacetylating histones and altering local chromatin structure (Jones et al. 1998; Nan et al. 1998), resulting in a more repressive environment.

DNA methylation clearly plays a role in transgene silencing, but whether it is involved in initiating silencing or maintaining the silent state is still unclear, though the evidence points to the latter. In their study, Mutskov and Felsenfeld (2004) demonstrate that DNA methylation is secondary to transgene silencing and suggest

that histone deacetylation may be the triggering event. In addition, Spencer and colleagues (2015) demonstrated that histone modifications appear to be responsible for the heterogeneity observed in transgene expression between clonal lines. However, Feng et al. (2001) showed that orientation dependent transgene silencing resulted from DNA methylation and was not due to changes in chromatin structure, determined by DNase I hypersensitivity; however, this does not rule out a role for histone modification. Interestingly, it has been shown that DNA methylation occurs at sites of DNA damage, such as that induced by site-specific nucleases (SSNs), but can be altered in a transcription-dependent manner (Morano et al. 2014). This evidence seems to indicate that DNA methylation, whilst clearly playing a role in transgene silencing, is not the initiating factor, though further studies are required to confirm this.

1.2.3 Transgene Properties

In addition to the local environment of the integration site, features within the transgene can also influence the likelihood of silencing. However, some elements have been found to reduce the likelihood of transgene silencing – notably insulators and ubiquitous chromatin opening elements (UCOEs).

1.2.3.1 Exogenous Promoters

Promoter choice is important to ensure efficient transgene expression. Viral promoters have commonly been used due to their high level constitutive expression and generally compact nature. However, eukaryotic cells have evolved mechanisms to detect viral promoters, such that many virally derived promoters are subject to silencing (Papadakis et al. 2004). One of the most commonly utilised viral promoters is the cytomegalovirus (CMV) promoter, however this often becomes silenced upon integration into the genome (Chen et al. 2011; Krishnan et al. 2006; Teschendorf et al. 2002).

In addition, eukaryotic promoters can be used to drive transgene expression. These have the advantage over virally derived promoters that they are not recognised as foreign. One of the most widely used constitutively active eukaryotic promoters is the elongation factor 1α (EF1α) promoter, which has been shown to be less prone to transgene silencing than the CMV promoter (Liu et al. 2009).

In addition to constitutively active promoters, tissue-specific promoter can be employed to restrict expression. This is particularly beneficial in a gene therapy setting as expression is restricted to the tissue of interest, so they have an improved safety profile and a more natural expression pattern.

1.2.3.2 Viral Elements

Retrovirally integrated transgenes can be silenced because of elements present in the integrated provirus. Some retroviral vectors contain silencer elements within their long terminal repeats (LTRs), which allow binding of trans-acting factors leading to the direct or indirect recruitment of chromatin remodelling factors and subsequent silencing of transgene expression (Pannell & Ellis 2001). These mechanisms may have evolved as a defence against viral infections. In their study, McInerney et al. (2000) demonstrated that long-term silencing of retroviral vectors cannot be reversed by employing DNMT or HDAC inhibitors, suggesting that mechanisms other than DNA methylation and histone deacetylation are involved. In order to reduce silencing of retroviral elements, insulators can be included in the viral 3'-LTR which can help block spread of repressive chromatin (Yi et al., 2011; discussed below).

1.2.3.3 Insulators

Chromatin insulators are genomic regions, which help to define differentially regulated loci, by blocking the action of *cis* regulatory elements. The best characterised insulator is derived from the chicken β-globin locus control region and contains a DNase I hypersensitive site (cHS4; chicken hypersensitive site 4; Emery et al., 2000). The cHS4 element is 1.2-Kb and is capable of maintaining expression of transgenes in a range of cell lines and model systems, though much of this activity is contained within a 250-bp region containing a 49-bp cHS4 core that interacts with CTCF (Emery et al. 2000).

Mutskov et al. (2002) showed that cHS4 is able to prevent CpG methylation within the promoter region and is associated with higher levels of histone acetylation, potentially via the recruitment of histone acetyltransferases (HATs) or inhibition of HDACs. While Emery et al. (2000) showed that cHS4 can help protect retroviral vectors from PEs. In addition, Sharma et al. (2012) showed that cHS4 can be used in

conjunction with transposon-based transgenes to prevent initial silencing. However, they show that cHS4 cannot completely protect against silencing of transgenes when multiple integrations occur.

1.2.3.4 Ubiquitous Chromatin Opening Elements (UCOEs)

UCOEs are genomic regions generally consisting of two divergently transcribed promoters, spanning a methylation free CpG island, capable of exerting a dominant chromatin remodelling function in order to maintain an open transcriptionally permissive local chromatin environment (Ackermann et al. 2014). The best characterised UCOE is derived from the human heterogeneous nuclear riboprotein A2 B1/ chromobox protein homologue 3 (HNRPA2B1-CBX3) housekeeping gene locus, termed A2UCOE (Antoniou et al. 2003). It has been successfully used to prevent transgene silencing of both gammaretro- and lentiviral gene therapy vectors in Chinese hamster ovary (CHO) cells, the hypermethylating cell line P19, as well as murine induced pluripotent stem cells (iPSCs) and embryonic stem (ES) cells (Williams et al. 2005; Zhang et al. 2010; Pfaff et al. 2013). In their study, Zhang et al. (2010) demonstrated that this is, at least in part, due to methylation resistance.

However, as the A2UCOE element extends into the first intron of either HNRPA2B1 or CBX3 or both, it has the potential for insertional mutagenesis as a result of aberrant splicing. Indeed, it has been shown to cause upregulation of neighbouring genes following aberrant splicing from the UCOE splice donor sites (Knight et al. 2012). There are two splice donor sites within A2UCOE plus additional potential cryptic splice donor sites that can lead to aberrant splicing. However, Knight et al. (2012) demonstrated that selective mutagenesis of these sites could alleviate aberrant splicing without compromising its ability to regulate silencing.

More recently, Müller-Kuller et al. (2015) have further improved the safety profile of A2UCOE by removal of the HNRPA2B1 promoter. The resulting UCOE – CBX3-UCOE – maintains the ability to provide an open chromatin environment with characteristic changes in promoter CpG methylation and histone modifications, such as reduced levels of repressive marks including H3K9me3 and H3K27me3. In addition, the activity of this minimal UCOE is locally restricted, such that it doesn't

override tissue specific promoters, adding greater specificity for use in a gene therapy setting.

Transgene silencing is a complex process involving many mechanisms, as discussed above. One of the key variables in determining transgene expression is the genomic environment at the integration site. However, it is important to bear in mind all potential mediators of silencing when creating a strategy for transgene integration.

1.3 Strategies and Tools for Transgene integration

Methods for transgene integration can be divided broadly into two groups: random integration and targeted integration.

1.3.1 Random Integration Strategies

Random integration of transgenes can be achieved by a variety of methods but all result in integration of transgenes in a random (or semi random) fashion and does not rely on the presence of homology between the targeting vector and the insertion site. Random integration mainly occurs as a result of the non-homologous end joining (NHEJ) pathway (Sakurai et al. 2010), though other pathways have also been implicated; notably, microhomology-dependent random integration (Merrihew et al. 1996) and the alternative end joining pathway (Iiizumi et al. 2008).

1.3.1.1 Plasmid DNA

Since the discovery 35 years ago that isolated DNA could stably integrate into the mouse genome (Gordon et al. 1980), this technique has become widespread in the generation of transgenic animals. This techniques leads to variable transgene copy number and the site of integration is unpredictable (Yan et al. 2010). The exact mechanism by which DNA becomes incorporated is not known, but is believed that microhomologies (1-10-bp) between the target site and transgene play a role via a process known as homologous illegitimate random integration (HIRI) (Yan et al. 2013; Yan et al. 2010). The advantage of this technique is the simplicity of targeting, the plasmid DNA does not require additional reengineering or preparation work prior to delivery (Fontes & Lakshmipathy 2013). However, integration efficiency is relatively low and so lengthy selection is required, generally by selection of drug resistance genes within the targeting cassette. In addition, it has been shown that up

to 10% of integrated transgenes carry aberrations – mainly deletions – which could hamper downstream applications (Recillas-Targa 2004).

1.3.1.2 Artificial Chromosomes

Artificial chromosomes, either bacterial, yeast or P1-derived (BACS, YACs and PACs, respectively) are incorporated into the host genome in an analogous manner to plasmid DNA. However, unlike plasmid DNA, BACs, YACS and PACs have a much greater cargo capacity in the range of 100-1000-Kb. BACS are the most widely used of these technologies and have been commonly used in the generation of transgenic mice (Chandler et al. 2007). Due to the large capacity of BACs, they can accommodate more of the endogenous regulatory landscape of the transgene, as such are often less prone to position effects and silencing than other randomly integrated transgenes (Beil et al. 2012; Giraldo & Montoliu 2001; Gong et al. 2003). However, this does not overcome issues associated with multiple integrations or adverse effects due to disruption at the site of integration. Indeed, in their study, Chandler et al. (2007), demonstrated that the majority of animals contained between 1 and 25 copies per genome. In addition, they showed a strong correlation between copy number and expression.

To overcome the issues associated with random integration and multiple copy integration, BACs can be used in conjunction with transposons to give more predictable outcomes. In their study, Suster et al. (2009) used BACs in conjunction with the Tol2 transposon for delivery into mice and zebrafish and were able to demonstrate singe copy integrations free from rearrangements or mutations and whose locations could be easily and precisely mapped.

1.3.1.3 Transposons

DNA transposons are class II transposable elements, which have been harnessed as tools for genome manipulation. Two notable examples are the Sleeping Beauty (SB) transposon and the PiggyBac (PB) transposon which are both widely used. The SB transposon is a synthetic transposon engineered from inactive members of the Tc1/mariner superfamily of transposons (Ivics et al. 1997). The PB transposon derived from *Trichoplusia ni*, has been successfully used in the creation of iPSCs and

is often used to deliver transgenes into a wide range of cell types from many organisms including mammals (Matteo et al. 2012).

DNA transposons consist of two functional elements: terminal inverted repeat (IR) sequences, which flank the gene to be mobilised and the transposase enzyme, which catalyses the reaction (Ammar et al. 2012). The transposase reaction proceeds in a "cut and paste" manner in which the transposase recognises the terminal IR sequences and catalyses the excision and reintegration of the transposon (Matteo et al. 2012). Integration of transposons is essentially random, but a minimal nucleotide sequence that varies between transposons is often required, for SB this is 5'-TA (Liu et al. 2005), while PB requires a 5'-TTAA sequence (Matteo et al. 2012).

1.3.1.4 Retroviral Vectors

Another method that improves the efficiency of transgene integration is the use of viruses, in particular retroviruses (Fontes & Lakshmipathy 2013). By their nature, retroviruses reverse transcribe their genome and integrate into the host genome via the action of the virally encoded integrase (Nisole & Saïb 2004). However, this integration is not entirely random, and different viruses show differing integration profiles; lentiviruses tend to integrate into active genes whilst gammaretroviruses prefer integration close to transcriptional start sites (Mitchell et al. 2004). One of the major concerns for the use of retroviral vectors in a gene therapy setting is their potential for oncogenesis. Indeed, following the successful treatment of X-linked severe combined immunodeficiency disorder (SCID-X1), four out of nine patients developed T-cell leukaemia due to the insertion of the gammaretroviral vector in close proximity to the *LIM-domain only 2* (LMO2) proto-oncogene, leading to its activation (Hacein-Bey-Abina et al. 2008). However, much work has focused on the improvement in the safety of viral vectors for use in gene therapy. This includes the use of self-inactivating (SIN) vectors lacking viral promoters and enhancers in their LTRs, to prevent promoting transcription of neighbouring genes. However, residual promoter activity still remains in some SIN vectors (Xu et al. 2012). Other groups have focused on reengineering integration site preference away from potentially harmful chromosomal locations. For example, El Ashkar et al. (2014) reengineered the viral integrase from the murine leukaemia virus (MLV) so that it could no longer

interact with bromodomain and extra-terminal (BET) family of proteins, normally responsible for directing integration. This vector has an altered integration profile, which is less prone to correlate with oncogene transcriptional start sites than the wild type vector.

1.3.2 Targeted Integration Strategies

Unlike random integration strategies, targeted strategies rely on the presence of homology between the targeting vector and the integration site; integration occurs in a predicable manner at a preselected target locus. This can help to alleviate some of the issues associated with random integration, such as insertional mutagenesis, copy number variation and the risk of gene silencing.

1.3.2.1 Homologous Recombination

Homologous recombination (HR) methods of transgene insertion rely on the use of the cell's HR machinery, and require the presence of regions of homology between the incoming DNA and the target locus. However, the efficiency of HR in somatic cells is relatively poor, though for unknown reasons, HR occurs more frequently in pluripotent ES cells (Houdebine 2002). Due to the relative rarity of HR, cells in which this occurs must be selected for, generally by both positive and negative selection, to avoid random insertion events. ES cells generated in this way can be injected into blastocysts or morula to generate chimeric embryos (Viville 1997). Whilst this method can be laborious, it has been successfully used in the generation of many transgenic mice.

1.3.2.2 ΦC31 Integrase

Recombinases are naturally occurring enzymes, which catalyse the integration of small fragments of DNA between two longer DNA fragments. The benefit of this methodology is that recombinases do not rely on the cell's HR repair machinery to deliver the transgene and as such it is theoretically applicable to any cell type, even quiescent cells in which HR does not occur.

ΦC31 integrase is a serine recombinase derived from the *Streptomyces* phage. It catalyses the insertion of the phage genome into the bacterial host genome via recombination between the phage attachment (attP) site and the bacterial attachment

(attB) site (Silva et al. 2011). Pseudo integration sites are found in mammalian genomes and have been exploited to induce integrase-mediated insertion events at specific loci in mammalian chromosomes. The most common site of integration in the human genome is on chromosome 19, position 19q13.31 (Chalberg et al. 2006), but approximately 100 other pseudo integration sites have been identified. A major disadvantage of ΦC31 is that integration can only occur at pseudo integration sites, thus cannot be targeted to predetermined loci without first engineering an integration site at the locus of interest.

To overcome the issues of semi-random integration more targeted approaches were developed, namely recombinase-mediated cassette exchange (RMCE), in which ΦC31 is used to drive integration of transgenes into an engineered locus containing the required recognition site. One disadvantage of this approach is the necessity to first generate a genomic target site harbouring the required recognition sequences. However, this system has been successfully used in a number of cell lines and model systems (Bateman et al. 2006; Chen et al. 2011; Belteki et al. 2003).

In addition, ΦC31 has been successfully used in proof of principle pre-clinical studies demonstrating its potential for gene therapy. For example, Ishikawa et al. (2006) showed that ΦC31 could drive integration of the common cytokine gamma chain gene in human T cell lines, resulting in stable expression and normal IL-2 signalling. This demonstrated its potential as an alternative strategy for treating SCID-X1.

1.3.2.3 Adeno-associated Virus (AAV)

AAV is a parvovirus with a linear ssDNA genome of 4.7-Kb, first identified as a "virus-like" particle in an adenovirus preparation (Hoggan et al. 1966). The discovery that low levels of the viral genome are integrated into the host cell genome (Cheung et al. 1980), led to the establishment of AAV as a potential vector for transgenesis, particularly in a gene therapy setting (Hastie & Samulski 2015). It has subsequently been shown that AAV can be used site-specifically to introduce both small and large alterations via HR with efficiencies of integration roughly 1% (Hirata et al. 2002). HR via AAV, whilst several orders of magnitude more efficient than HR alone, still requires the presence of large regions of homology, generally several

kilobases long (Khan et al. 2011), which can be challenging to construct. However, AAV can target a wide range of cell types and has been used successfully to target immortalised cell lines, primary cells and stem cells – both ES cells and iPSCs (Khan et al. 2011).

One of the advantages of AAV-based vectors is their ability to be pseudotyped to specifically target tissues of interest. For example, AAV4 is tropic for the lungs, while AAV9 shows a greater preference for the liver and heart (Asokan et al. 2012; Zincarelli et al. 2008).

1.3.3 Endonucleases

Endonuclease-mediated methods of genome engineering rely on the formation of targeted DNA double-strand breaks (DSBs), which increase the rate of HR by several orders of magnitude in mammalian cells (Elliott et al. 1998; Liang et al. 1996; Choulika et al. 1995; Rouet et al. 1994b). One of two cellular pathways repairs these DSBs: HR repair or NHEJ repair. HR is an error-free pathway which uses a template to faithfully repair the DSB, as such can be harnessed to create targeted gene insertions or to correct gene mutations. However, HR repair is most active during late S and G2 phases of the cell cycle (Fung & Weinstock 2011) and does not occur in quiescent cells. In contrast, NHEJ predominates in G0 and G1 phases, but remains active throughout the cell cycle (Fung & Weinstock 2011); generally, NHEJ is more active than HR (Sakurai et al. 2010). NHEJ is an error-prone pathway leading to small insertions or deletions (indels) at the site of the DSB so can be harnessed to create gene knockouts.

1.3.3.1 Meganucleases

Meganucleases (also known as homing endonucleases) have been used for over 15 years in genome editing applications (Silva et al. 2011). They are found in all kingdoms of life and are generally encoded within introns or inteins (Silva et al. 2011). They are highly specific and in nature have single target loci within their host genomes (Stoddard, 2011). Meganucleases have been categorised into five families based on sequence and structural motifs: LAGLIDADG, GIY-YIG, HNH, His-Cys box and PD-(D/E) XK.

The LAGLIDADG family are the most well studied and are commonly used in gene targeting applications. All members of this family contain the defining LAGLIDADG amino acid sequence which is an essential element for their enzymatic activity (Silva et al. 2011). They specifically recognise a 14-40-bp DNA sequence and cleavage results in a characteristic 4-nt 3'-OH overhang across the minor groove of DNA (Silva et al. 2011).

DNA-binding specificity of meganucleases shows considerable context dependence influenced by neighbouring amino acids and DNA-contacts (Schiffer et al. 2012). Despite this, the DNA-binding specificity has been successfully reengineered, predominantly using the I-CreI scaffold (a member of the LAGLIDADG family), derived from *Chlamydomonas reinhardtii*.

In addition to endonuclease activity, meganucleases have been reengineered to produce DNA single-strand breaks (nicks); such meganucleases are termed meganickases. These are capable of stimulating HR (though with reduced efficiency compared to meganucleases) but do not stimulate the error-prone NHEJ pathway (Schiffer et al. 2012).

1.3.3.2 Zinc Finger Nucleases (ZFNs)

ZFNs are composed of 3-4 zinc finger (ZnF) domains fused to the non-specific nuclease domain of the FokI endonuclease via a flexible linker (Figure 1.2). FokI is a type IIS restriction endonuclease first isolated from *Flavobacterium okeanokoites*. Unlike type II restriction endonucleases, commonly used in molecular biology, type IIS endonucleases cut outside of their DNA recognition site (Carroll 2014). The recognition and cleavage domains of FokI can be separated by limited proteolysis (Li et al. 1992) and early studies demonstrated that its recognition specificity could be altered by linking the nuclease domain to alternative DNA-binding domains, including ZnF domains (Kim et al. 1996). FokI nuclease is an obligate dimer: thus two ZFNs, binding opposite strands of DNA, are required for DSB formation. Following dimerisation of its catalytic domain, FokI cleaves DNA resulting in a DSB with characteristic 5' overhangs (Bitinaite et al., 1998).

ZnFs are one of the most common structural motifs in the eukaryotic proteome (Rubin et al. 2000). They are small peptides (30 amino acids), which contain a Cys2-His2 motif and form a characteristic ββα configuration (Pabo et al. 2001). The α-helix projects into the major groove of DNA where it contacts 3-4 contiguous bases (Perez-Pinera et al. 2012), such that each ZnF domain recognises a specific triplet of DNA bases. Synthetic ZnF domains, which can specifically recognise almost all 64 possible nucleotide triplets, have been generated by site directed mutagenesis, rational design, and selection of large combinatorial libraries (Perez-Pinera et al. 2012); thus ZFNs can theoretically be designed to target almost any genomic sequence. However, in practice, DNA-binding specificity shows context dependence largely influenced by neighbouring ZnF domains. Despite this, ZFNs have been successfully utilised for many genome editing applications including gene correction (Urnov et al. 2005), gene addition (Moehle et al. 2007) and gene knockouts (Perez et al. 2008). Indeed, ZFNs have shown promising results in initial clinical trials for human immunodeficiency virus 1 (HIV-1) positive patients. Early studies demonstrated the ability of ZFNs to knockout the gene encoding major co-receptor of HIV-1 viral entry CCR5 (chemokine [CC motif] receptor 5) in T-cells which resulted in lower viral loads following engraftment in mice (Perez et al. 2008). Subsequent human trials have demonstrated the promise of transplantation of ZFN modified autologous T-cells in the treatment of HIV infection (Tebas et al. 2014).

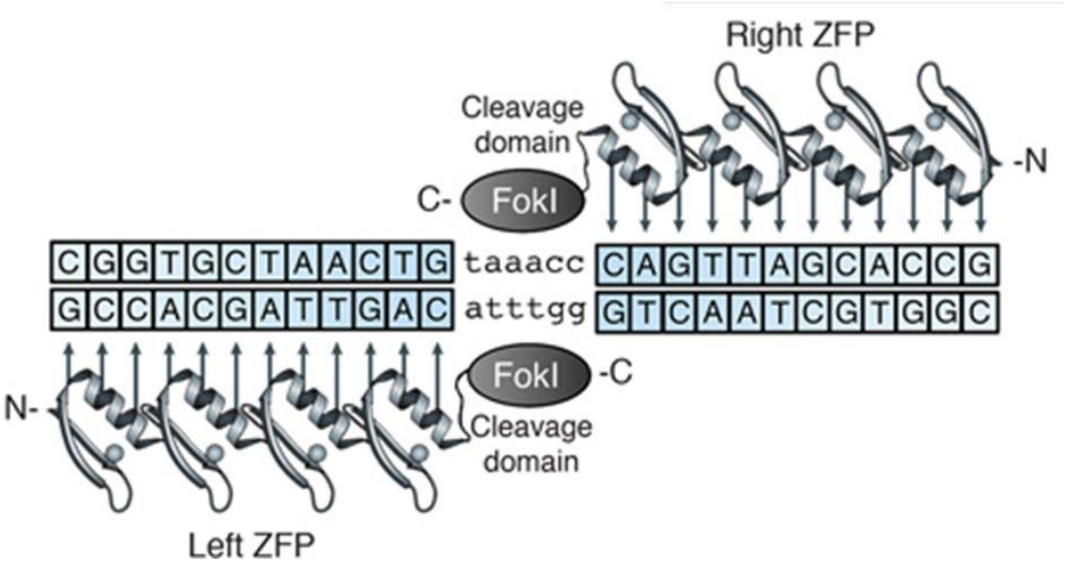


Figure 1.2 ZFNs binding a DNA target. Each ZFN consists of 3-4 ZnFs, capable of recognising a specific triplet of DNA bases (arrows). The ZnFs are conjugated to the non-specific nuclease domain of FokI endonuclease, via a flexible linker. Due to the requirement of FokI to dimerise, target sites, on opposite DNA strands must be separated by 5-6-bp (when a short linker is used). In the commonly used system, the FokI domain is conjugated to the C-terminus of the ZnF cluster. Reprinted by permission from Elsevier Ltd: Trends in Biotechnology (Gaj et al. 2013)**, copyright (2013).**

1.3.3.3 Transcription Activator-Like Effector Nucleases (TALENs)

TALENs are programmable, targeted nucleases formed by the fusion of a C-terminal truncation of a transcription activator-like effector (TALE) with the nuclease domain of the non-specific nuclease domain of FokI.

TALEs are naturally occurring proteins found in phytopathogenic bacteria of *Xanthomonas* species (Figure 1.3). In nature they are secreted via a type III secretion system into the host cell wherein they are trafficked to the nucleus, interact with host DNA, and promote transcription of genes beneficial to bacterial colonisation and spread (Boch & Bonas 2010).

In 2009 the TALE cipher ('TALE code') was elucidated, revealing that each monomer in the DNA-binding domain binds to a single DNA base (Figure 1.3; Boch et al. 2009; Moscou & Bogdanove 2009) allowing the generation of TALEs to target almost any genomic sequence. Each monomer is composed of 33-35 highly conserved amino acids. Two hypervariable residues found at positions 12 and 13, termed the repeat variable di-residue (RVD), determine DNA-binding specificity. Residue 13 contacts the DNA base while residue 12 stabilises the interaction (Bochtler 2012). The four most common RVDs (NI, HD, NG and NN) bind the four DNA bases (adenine, cytosine, thymine and guanine, respectively). However, NN can also bind to adenine in some contexts. Other monomers have been identified with alternative binding specificities, for example NK and NH, which have greater specificity for guanine, but are considered weaker than NN (Christian et al. 2012; Streubel et al. 2012; Cong et al. 2012; Meckler et al. 2013). Unlike ZFNs, each monomer repeat within a TALEN recognises and binds to a single DNA-base and to date no context dependence has been observed, meaning TALENs are much more flexible than ZFNs and can theoretically bind to any genomic sequence. However, the preferred base in the 0 position is a T, specified by a cryptic signal upstream of the monomer array, which is similar in structure but not sequence to the other monomers (Mak et al. 2012). In nature, TALEs have been found to contain between 1.5 and 33.5 monomers, though it has been demonstrated experimentally that arrays containing less than 6.5 repeats have limited activity (Boch et al. 2009). The final monomer is termed the half repeat (or 0.5 repeat) and is composed of 20 amino acids.

Like ZFNs, TALENs must work in pairs, binding opposite strands of DNA, to allow the FokI nucleases domains to dimerise (Figure 1.4). For successful DSB formation, the distance between each TALEN binding site, known as the spacer region, must be 14-20-bp (Sanjana et al. 2012). TALENs have been shown to be at least as efficient as ZFNs in side by side comparisons (Mussolino et al. 2011), with a recent study showing they can induce up to 10-fold more mutations, via NHEJ repair of DSBs, in a zebrafish model (Chen, Oikonomou, et al. 2013).

Certain criteria must be met for TALENs to successfully bind and cleave at a genomic locus of interest. Cermak and colleagues (2011) proposed five criteria based on investigation of naturally occurring TALEs. 1) The monomer binding site should be preceded by a 5'-T. 2) There should not be a T at position one. 3) There should not be an A at position two. 4) They should end with a T. 5) They should have a base composition within two standard deviations of the averages which Cermak et al. (2011) observed. In order to identify targets which meet these criteria within a genomic region of interest, Cermak and colleagues (2011) devised a web-based platform which will screen a given genomic region ≤5-Kb. Subsequently, four additional criteria were proposed in order to maximise TALEN activity: 1) include at least 3-4 strong RVDs; 2) position strong RVDs to avoid long (≥6) stretches of weak RVDs, especially at the ends; 3) for high guanine specificity use NH or NK (if enough strong RVDs are present), and 4) use NN for guanine if only a few other strong RVDs are present (Streubel et al. 2012).

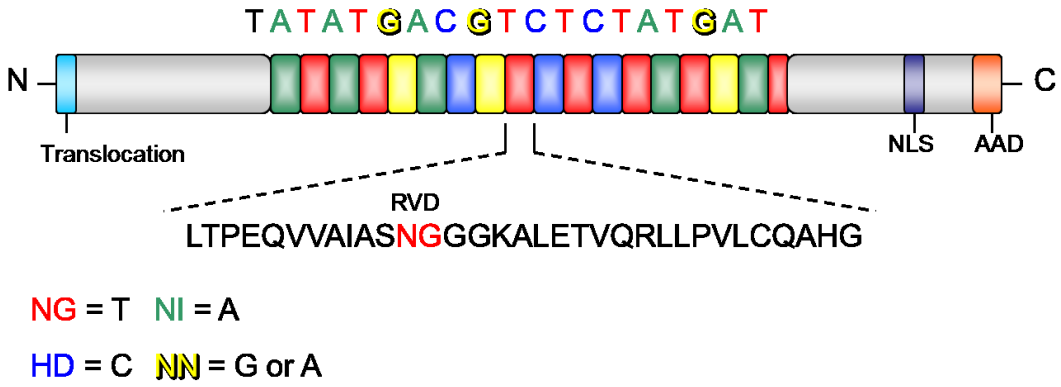


Figure 1.3 Structure of a TALE molecule. TALEs consist of three domains: the N-terminus, the C-terminus, and the DNA-binding array. The N-terminus contains a translocation domain (required for secretion) and a cryptic signal that recognises thymine as the first base of the DNA-binding site (Sanjana et al. 2012); the C-terminus contains a nuclear localisation signal (NLS) and an acidic activation domain (AAD) required for transcription factor function (Bogdanove et al. 2010). The DNA-binding array is made up of 1.5-33.5 monomer repeats or monomers. Each monomer recognises one base of DNA, with specificity conferred by the repeat variable di-residue (RVD). The four most common RVDs recognise one of the four bases of DNA: NG recognises thymine, NI recognises adenine, HD recognises cytosine, and NN recognises guanine or adenine. Figure adapted from Sanjana et al. (2012).

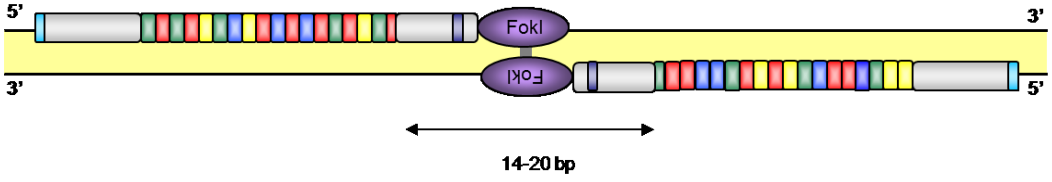


Figure 1.4 TALENs bound to target site. TALENs bind in pairs, binding opposite strands of DNA. The distance between the two TALEN target sites is crucial to allow the FokI nuclease catalytic domains to dimerise and must be 14-20 bp. Once activated, FokI cleaves the DNA on both strands resulting in a DSB. Unlike ZFNs, TALENs bind their targets in a forward orientation, so that cleavage occurs at the 3' end of each half site. Figure adapted from Sanjana et al. (2012).

The Zhang laboratory devised a method for TALEN construction (Sanjana et al. 2012), which uses a hierarchical ligation strategy and makes use of the powerful Golden Gate cloning technique (Engler et al., 2009, 2008; Weber et al., 2011; oulined in Figure 1.5). The first stage involves the production of a monomer library by PCR amplification of the four monomers (containing the four most common RVDs – NG, NI, HD, and NN) using primers with ligation adapters, resulting in 18 different monomers for each of the four templates (72 monomers in total). The ligation adapters are designed such that the position of each monomer within the array is uniquely defined. Once generated, this library can be used to construct arrays containing any combination of monomers, allowing the construction of TALENs containing any combination of monomers in their DNA-binding domains.

Following identification of an appropriate target site, the sequence of monomers is divided into three groups of six – termed hexamers. Monomers from the library are ligated into circular hexamers using the first Golden Gate cloning reaction (Figure 3.2, Stage 3). Treatment with PlasmidSafe DNase degrades any non-ligated products, as only correctly assembled hexamers can circularise.

Each hexamer is then amplified, gel purified and ligated into the appropriate backbone plasmid using the second Golden Gate cloning reaction. Constructed plasmids are then transformed in to competent *E. coli*. To assess the fidelity of the assembly, a colony PCR is carried out and positive colonies are picked for plasmid extraction, and sequencing.

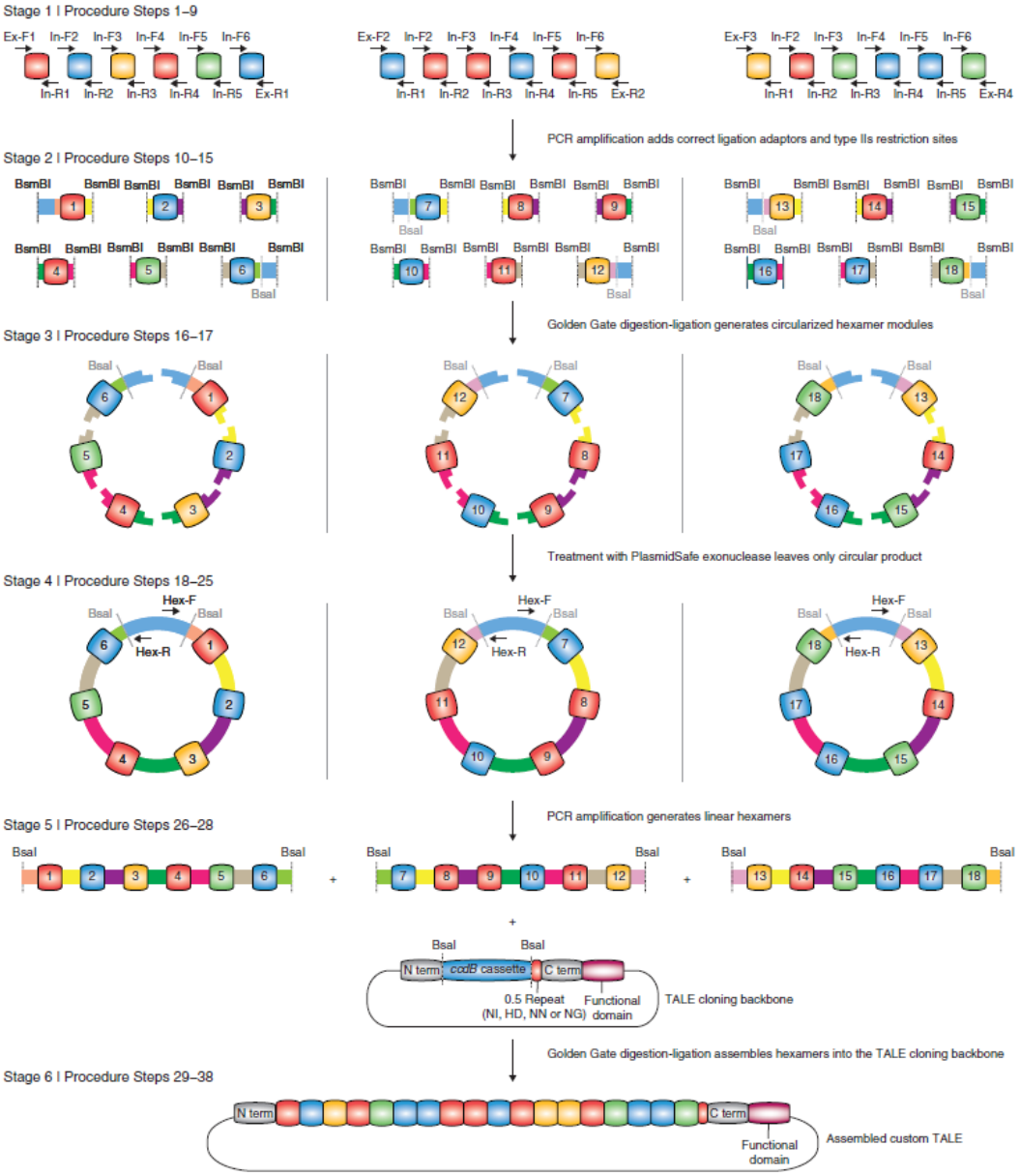


Figure 1.5 Construction of TALENs. The method devised by Sanjana et al. (2012) uses a hierarchical ligation strategy. Stage 1: Production of a library of monomers containing ligation adapters, which will specify their position in the monomer array. Stage 2: Individual monomers are ligated to form three hexamers (1-6, 7-12 and 13-18) using the Golden Gate cloning strategy. Stage 3: Each ligation adapter is designed so that the correctly ligated hexamers will form complete circles; any incorrectly ligated products are removed by the action of PlasmidSafe DNase. Stage 4: Correctly, ligated hexamers are individually amplified to create linear molecules. Stage 5: The three hexamers are ligated into an appropriate backbone vector (which encodes the N- and C-termini, including the cryptic T signal and 0.5 repeat) using the second Golden Gate reaction. Stage 6: Assembled TALENs are transformed into competent *E. coli*, successful clones are purified, and sequence verified. Reprinted by permission from Macmillan Publishers Ltd: Nature Protocols (Sanjana et al. 2012), copyright (2012).

TALENs appear to be very specific, with few if any off-target events reported (Guilinger, Pattanayak, et al. 2014; Hockemeyer et al. 2011; Mussolino et al. 2011) and side by side comparisons show reduced off-target induced cytotoxicity compared to ZFNs for some target sites (Mussolino et al. 2011).

TALENs have been successfully used in a range of tissues and in the generation of transgenic animals, including but not limited to zebrafish, mice, pigs and cows (Wright et al. 2014).

Building on the success of trials using ZFNs for treatment of HIV-1 infection (detailed above), recent studies have demonstrated the potential of TALENs in both the eradication of integrated provirus (Ebina et al. 2015; Strong et al. 2015) and also in the prevention of viral entry via CCR5 disruption (Mock et al. 2015; Mussolino et al. 2014).

In addition to nuclease activity, TALEs have the potential to be fused with an array of functional domains such as transcriptional activator domains (e.g. VP64), repressor domains and chromatin remodelling domains, e.g. cytidine deaminases, HATs and DNMTs (Bogdanove and Voytas, 2011) enabling highly specific genome editing and regulation.

1.3.3.4 Clustered, Regularly Interspaced Short Palindromic Repeats (CRISPR)/ CRISPR Associated 9 (Cas9)

The latest development in SSN technology is the CRISPR/ Cas9 system (Cong et al. 2013; Mali, Yang, et al. 2013; Jinek et al. 2012). Based on the prokaryotic type II CRISPR/Cas system, this technology has been adapted for use in mammalian (and other) model systems. The prokaryotic CRISPR/Cas systems act like adaptive immune systems in bacteria and archaea, identifying and targeting the degradation of invading nucleic acid species through short complementary RNAs expressed from the CRISPR locus. This locus is composed of short sequences, termed spacers, acquired from the integration of DNA from invading viruses or plasmids separated by direct repeats. In the type II system, these are expressed as long pre-CRISPR RNAs (pre-crRNAs), which are processed by association with trans-activating crRNAs (tracrRNAs), RNase III, and Cas9 to produce mature crRNAs. Mature

crRNAs interact with tracrRNAs via their repeat sequences (Jinek et al. 2012) and form a complex with Cas9. This recognises invading nucleic acid species, through complementarity between the crRNA and sequences within the invading nucleic acid, termed protospacers as they gave rise to the spacer sequence in a prior infection. This leads to the direct cleavage of the incoming DNA species via the action of the Cas9 nuclease. Cas9 contains two nuclease domains, homologous to HNH and RuvC endonucleases; in wild type Cas9, the HNH-like domain cleaves the complementary DNA strand whilst the RuvC-like domain cleaves the non-complementary DNA strand (Jinek et al. 2012). Cleavage of the target by Cas9 is dependent on both base-paring between the crRNA and the presence of a specific motif adjacent to the genomic target site known as the protospacer adjacent motif (PAM; Jinek et al., 2012).

The most widely used system for gene editing is based on that of *Streptococcus pyogenes.* It is composed of two main elements: the synthetic single guide RNA (sgRNA), and the Cas9 protein (Cong et al., 2013; Jinek et al., 2012; Mali et al., 2013c; Figure 1.6). The sgRNA is a fusion of the crRNA and the tracrRNA, obviating the need for RNase III, and is responsible for target recognition and binding via a 20-bp region homologous to the target DNA (protospacer). Interaction of the sgRNA with the Cas9 causes a conformational change in the Cas9 protein, which facilitates the recognition and binding of the target site in a PAM-dependent manner (Anders et al. 2014). Indeed, recognition of the PAM is a prerequisite for ATP-dependent strand separation and subsequent sgRNA-target DNA heteroduplex formation (Sternberg et al. 2014). The PAM motif for *S. pyogenes* Cas9 is 5'-NGG, which must be present directly 3' to the target; cleavage occurs 3-bp upstream of the PAM (Jinek et al. 2012).

Unlike ZFN or TALEN technologies, CRISPR/Cas9 does not require the reengineering of proteins in order to redirect its target specificity, meaning it is both quick and cost effective to produce sgRNAs targeting any loci of interest. Once target sites have been identified, oligonucleotides encoding the 20-bp sgRNA target sequence (and its reverse compliment) with 5' ligation adapters can be synthesised (Figure 1.7). These oligonucleotides are annealed and phosphorylated, then cloned

into the backbone vector encoding the sgRNA scaffold, using a Golden Gate reaction (Figure 1.7). Constructed plasmids are transformed into *E. coli*, isolated, purified and then sequence verified.

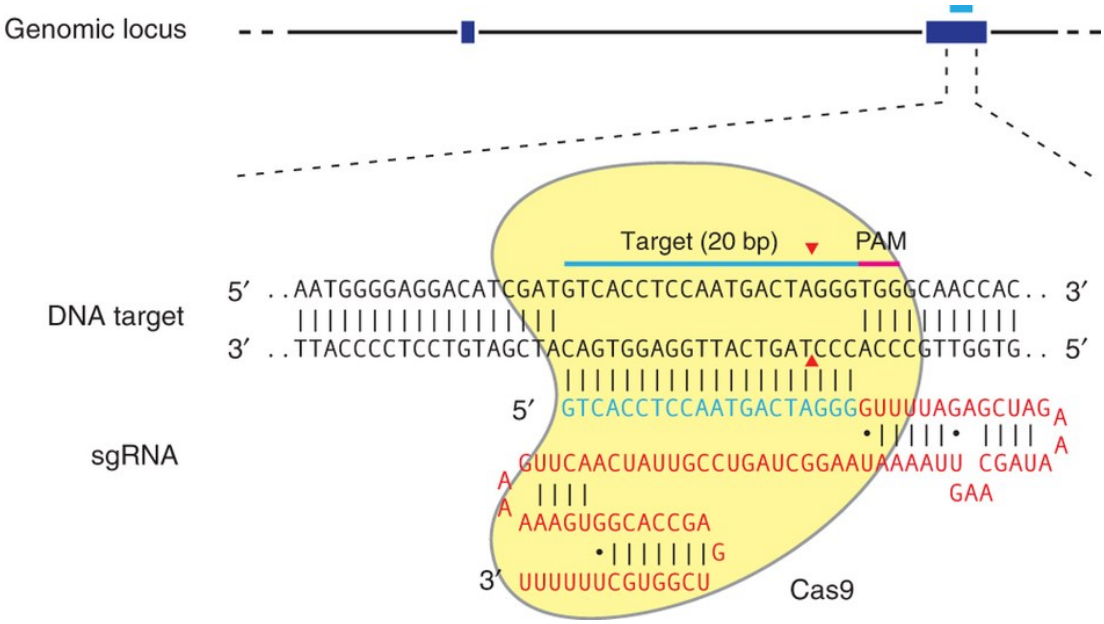


Figure 1.6 CRISPR/Cas9 target recognition and cleavage. The sgRNA binds to its genomic target via Watson-Crick base pairing, this allows Cas9 to bind and cleave the target sequence following recognition of the PAM in the form 5'-NGG for S. pyogenes. Cas9 cleaves the target locus approximately 3-bp upstream of the PAM, resulting in a blunt-ended DSB. Reprinted by permission from Macmillan Publishers Ltd: Nature Protocols (Ran, P D Hsu, et al. 2013), copyright (2013).

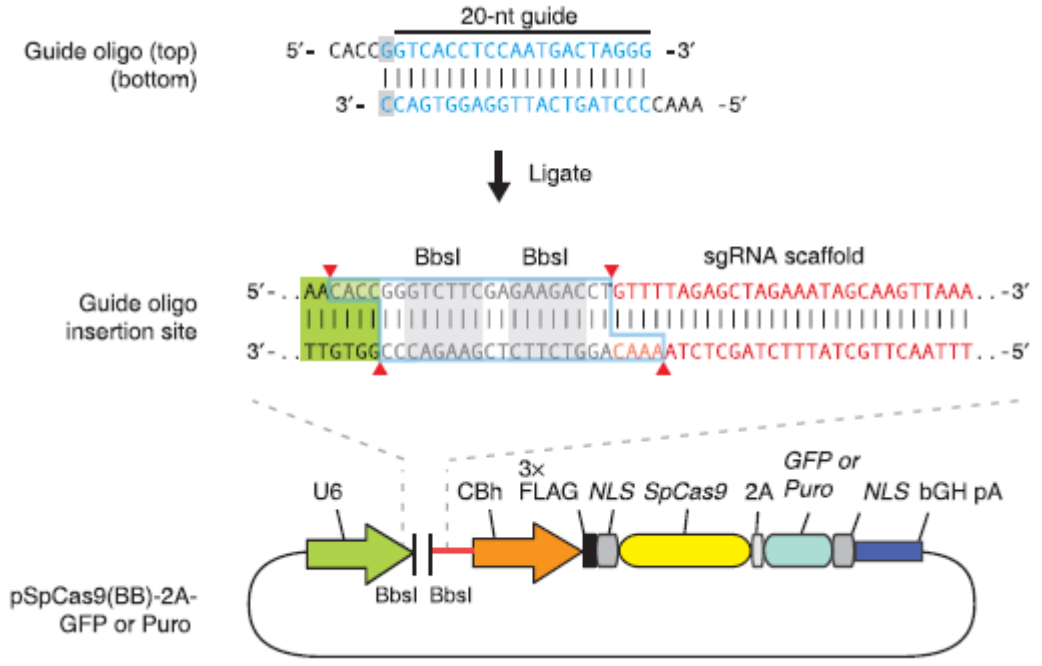


Figure 1.7 sgRNA construction using the Zhang group method (Ran, P D Hsu, et al. 2013). Oligonucleotides encoding the top and bottom of the 20-bp target (blue) with ligation adapters are synthesised. These are annealed and phosphorylated before ligation into the backbone plasmids using a Golden Gate cloning reaction. The backbone plasmid encodes the sgRNA scaffold and a human codon optimised Cas9 or Cas9n. In addition, a puromycin resistance gene allows for successfully transfected cells to be selected. Reprinted by permission from Macmillan Publishers Ltd: Nature Protocols (Ran, P D Hsu, et al. 2013), copyright (2013).

Shortly after its inception, it became apparent that relatively high frequency off-target cleavage could occur (Fu et al. 2013; Hsu et al. 2013; Mali, Aach, et al. 2013; Pattanayak et al. 2013). These studies into the fidelity of the RNA-DNA interaction revealed that even in the presence of up to five base pair mismatches, the sgRNAs could target Cas9 activity, and activity at these sites could be equal to or greater than that observed for the intended target site (Fu et al. 2013). This is particularly apparent when these mismatches are found in the 5' portion of the sgRNA – distal to the PAM sequence. Indeed, it was also shown that a PAM sequence of 5'-NAG could also support Cas9 activity. This was later confirmed by structural analysis of the Cas9 protein (Anders et al. 2014). In addition, each of these studies demonstrated that off-target activity appeared to be sgRNA dependent; individual guides could tolerate different mismatches (both number and location). Most of these studies indicated that this off target activity was linked to Cas9 concentration; stringency of target recognition is reduced when Cas9 concentration is increased (Hsu et al. 2013; Mali, Aach, et al. 2013; Pattanayak et al. 2013). This ability to recognise mismatches may be beneficial in the bacterial CRISPR/Cas systems, which are designed to target invading viral species that have rapidly evolving genomes; the ability to recognise similar but not exactly matching protospacers increases the repertoire of sequences that can be targeted.

Different strategies were developed to address this issue. The double nickase strategy makes use of a mutant Cas9 protein: Cas9 nickase (Cas9n). Cas9n carries the D10A mutation, which renders its RuvC-like catalytic domain inactive; it induces single strand breaks – nicks – rather than DSBs (Figure 1.8). These nicks can be used to drive HR repair alone, or two sgRNAs targeting opposite strands of the DNA (between 0-pb and 30-bp between 5' ends) can be used to form DSBs with 5' overhangs (Figure 1.8; Mali et al., 2013a). If nicks are induced at off target sites, they are less mutagenic than DSBs as they are repaired via the base excision repair pathway, which like HR is much less error-prone than NHEJ.

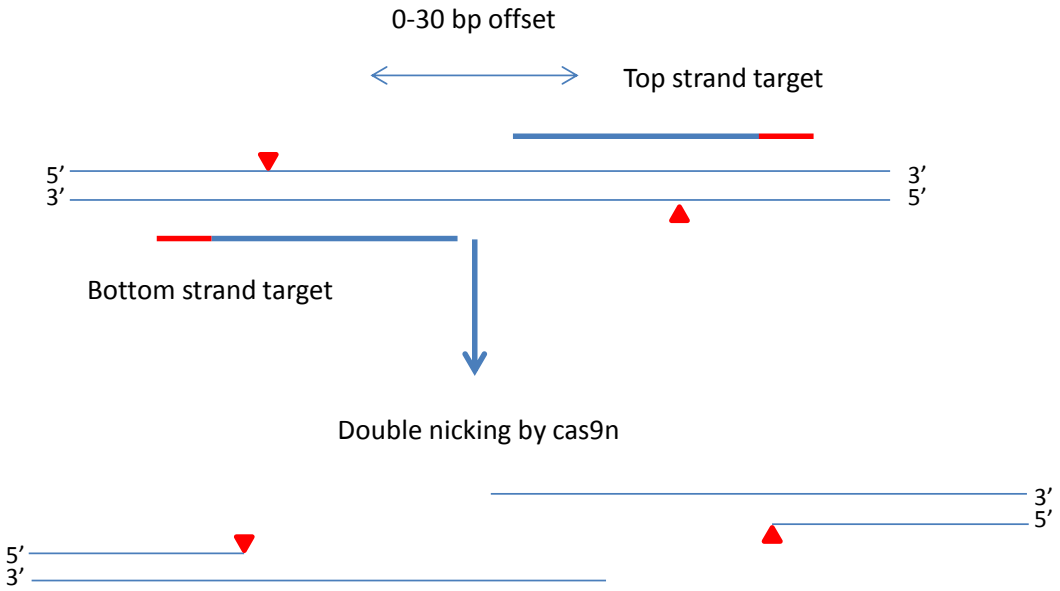


Figure 1.8 CRISPR/Cas9 double nickase strategy. Red lines indicate PAM sequence. In this case, targets are located in a tail-to-tail orientation which has been shown to be more efficient than a head-to-head orientation (Shen et al., 2014). This method is predicted to be far less mutagenic as off-target nicks have less mutagenic potential than off-target DSBs. In addition, the requirement for two sgRNAs reduces the chances of DSBs occurring off-target. Figure adapted from Ran et al. (2013).

The orientation of the sgRNAs in relation to each other seems to affect the cleavage efficiency. Mali and colleagues (2013) indicated that 5' overhangs are more efficient than 3' overhangs at inducing an NHEJ repair response. However, Shen et al. (2014) showed equally robust NHEJ resulting from either 3' or 5' overhangs (generated using Cas9 nucleases deficient in either the HNH domain or the RuvC-like domain). However, they also demonstrated that tail-to-tail orientation is more efficient than head-to-head orientation at inducing indels. Though less frequent, Cas9n induced SSBs can still result in indels at off-target sites, indicating that some SSBs can be converted into DSBs and are repaired via NHEJ (Cho et al. 2014).

In addition to the use of paired nickases, it was suggested that the use of shorter sgRNAs would also increase the stringency of binding site recognition as they would be more sensitive to mismatches, thus lead to reduced off target activity (Fu et al. 2014). In their study, Fu and colleagues (2014) were able to demonstrate that sgRNAs containing 17-bp of homology could efficiently induce on-target indels and drive HR, but showed a better off target profile than their full length counterparts. Though again, sensitivity to mismatches, and thus ability to induce off-target events was guide dependent.

More recently, FokI-Cas9 fusions have been utilised to further increase the fidelity of on-target activity and reduce potential off-target recognition (Guilinger, Thompson, et al. 2014; Tsai et al. 2014). This method makes use of the nuclease dead Cas9, which is fused to FokI. Due to the nature of FokI, nuclease activity can only occur when targets sites are positioned to allow for FokI dimerisation, as is the case for ZFNs and TALENs. This means that even if the sgRNAs bind to off-target sites, the FokI-Cas9 will not be active, unless an adjacent site allows for binding of another sgRNA. Indeed, Tsai and colleagues (2014) demonstrated that in a direct comparison, monomeric Cas9 nickase could induce indels at a much higher frequency than monomeric FokI-Cas9. In addition, they showed that monomeric Cas9 nickases could induce point mutations with relatively high frequency (Tsai et al. 2014).

In a direct comparison, the CRISPR/Cas9 system showed more off-target activity than TALENs, but this was much lower than initial studies implied, and no off-target

activity was detected when a nickase Cas9 was used (Wang et al. 2015). Generally, studies investigating off-target activity suggest that this is less of an issue than first believed, especially if careful sgRNA design is implemented (Cho et al. 2014; Mandal et al. 2014; Smith et al. 2014; Suzuki et al. 2014; Veres et al. 2014).

One of the benefits of the CRISPR/Cas9 system is the small size of the sgRNAs, making it amenable to multiplexing. This can be used for the deletion of large segments, by creating two DSBs either side of the region to be deleted (Canver et al. 2014; Ran, Patrick D Hsu, et al. 2013; Ran, P D Hsu, et al. 2013; J. Zhou et al. 2014; Xiao et al. 2013; Gupta et al. 2013), though this has also been achieved with ZFNs (Lee et al. 2010). It can also be used to create precise genome rearrangements – inversions and translocations – to mimic those observed in diseases such as cancer (Choi & Meyerson 2014). In addition, this has the potential to target multiple genes simultaneously, which could streamline the production of transgenic animals with complex genotypes (Yang et al. 2013).

Like the TALEN technology, CRISPR/Cas9 has been utilised for genomic manipulations other than nuclease activity, use of nuclease dead Cas9 conjugated to other enzymes, in particular VP64 for use as a transcriptional activator (Mali, Aach, et al. 2013), or repressors (Gilbert et al. 2013). Though other uses, including chromatin remodelling, modulation of genomic architecture and Cas9-guided recombinases have been proposed (Mali, Esvelt, et al. 2013).

1.3.3.5 SSN-Mediated Homology Directed Repair (HDR)

One of the key advantages of the SSNs is their ability to increase the efficiency and specificity of transgene integration via homology directed repair (HDR). Early studies demonstrated that the provision of a DNA DSB can increase the efficiency of transgene integration by several orders of magnitude (Choulika et al. 1995; Elliott et al. 1998; Liang et al. 1996; Rouet et al. 1994b). While ZFNs, TALENs, and CRISPR/Cas9 each have distinct methods for recognising, binding and cleaving their target sites (discussed above) each can be used to drive HDR by the provision of a repair cassette with appropriate homology to the target locus. This can be used for either gene repair or transgene insertion (Figure 1.9).

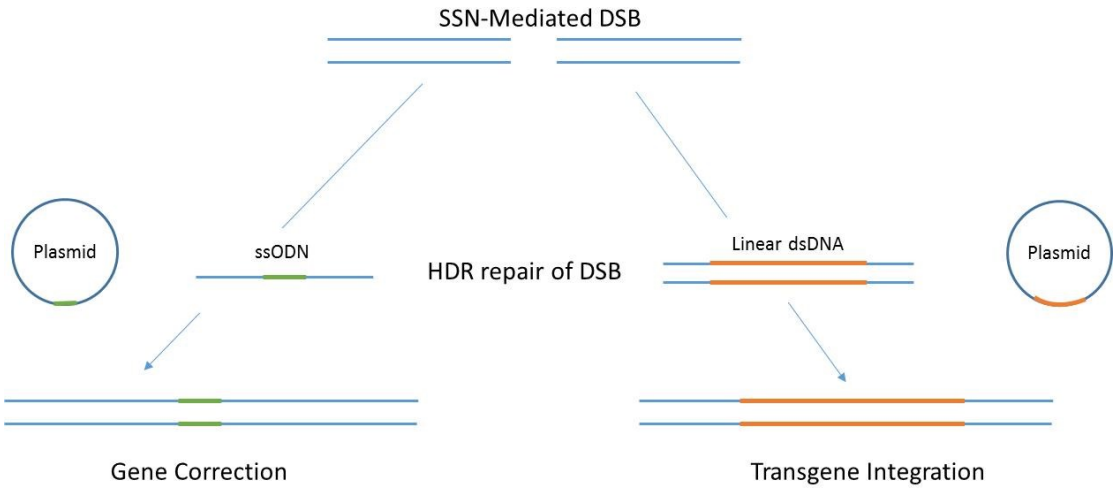


Figure 1.9 SSN-mediated HDR. Targeted DSBs generated following activity of SSNs can be repaired using HDR if a repair template is provided. To make small changes, such as single base changes (often referred to as gene editing), the repair template can either be double stranded or single stranded. For larger changes, such as gene insertions, a dsDNA template is used (either plasmid or liner).

In order for HDR to occur, a repair template must be provided, that can either be single stranded DNA (ssDNA) generally in the form of an oligonucleotide (ODN) or double stranded DNA (dsDNA), either in linear or circular form (plasmid).

In their study, Urnov et al. (2005) used ZFNs to generate a single point mutation in the interleukin-2 receptor gamma chain (*IL2Rγ*) gene via HDR using a plasmid donor. They were able to achieve an HDR frequency of 18%, with 7% of events occurring biallelically. In addition, ZFNs have also been used to successfully integrate larger DNA fragments site-specifically. In their study, Moehle and colleagues (2007) were able to demonstrate HDR efficiencies of 5%, 6% and 15% for integration of a 1.5-Kb promoter-transcription unit, a 900-bp open reading frame, and a 12-bp tag, respectively.

The development of TALENs greatly facilitated the ability to correct gene mutations and drive transgene integration, due to their relative ease of design and construction compared with ZFNs. This accelerated the use of gene editing particularly for potential therapeutic applications. For example, Sun et al. (2012) demonstrated that TALENs could be used to specifically edit the human β-globin (*HBB*) gene, associated with sickle cell disease. They were able to demonstrated that TALENs increased the editing efficiency >1000-fold and showed no detectable cytotoxicity. In addition, TALENs have been used in proof of principal experiments in patient-derived iPSCs from SCID-X1 patients. Edited cells were capable of producing haematopoietic precursors as well as mature NK cells and T cell precursors expressing the correctly spliced IL-2Rγ (Menon et al., 2015).

The advent of the synthetic CRISPR/Cas9 system further accelerated the use of gene editing. In their initial study, Mali et al. (2013) were able to demonstrate the use of this system for both small gene correction using an ssODN template and insertion of larger transgenes from plasmid vectors, with efficiencies greater than comparable TALENs. Subsequent studies have confirmed the efficiency of this system; Staahl and Doudna were able to demonstrate CRISPR/Cas9 mediated HDR rates of up to 38% when the system components were delivered as ribonucleoprotein complexes. In addition, this system has greatly facilitated the production of transgenic animals

from a range of species (Quadros et al. 2015; Auer et al. 2014; Chen, Fenk, et al. 2013).

To further improve the efficiency of HDR, Yu et al. (2015) employed the use of small molecules to bias the outcome of DSB repair in favour of HDR over NHEJ. Using L755507 (a β3 adrenergic receptor partial agonist) and Brefeldin A (an inhibitor of retrograde transport from the endoplasmic reticulum to the Golgi apparatus), they were able to demonstrate efficiencies 3-fold higher for large insertions and up to 9-fold higher for point mutations. The mechanisms via which they increase the efficiency of HDR are yet to be elucidated.

1.4 Target Sites for Long-Term Expression

Targeted transgene insertion approaches rely on the selection of permissive loci, in order to achieve robust long-term expression of the transgene. As discussed above (Section 1.2.1), the local chromatin environment of the integrated transgene influences whether it will become silenced. In order to reduce the risk of transgene silencing, it is important to identify loci which can support transgene expression. It is generally understood that open chromatin is more easily targeted than heterochromatin, particularly by SSNs (van Rensburg et al. 2013). However, transgenes inserted by random integration strategies are more prone to PEs due to proximity to heterochromatin. In this work, I have investigated the potential of four human genomic loci to support robust long-term transgene expression: the adeno-associated virus (AAV) integration site 1 (AAVS1); the human homologue of the mouse *Rosa26* locus (hROSA26); the inosine monophosphate dehydrogenase 2 (*IMPDH2*) gene and the eukaryotic translation elongation factor 1 alpha 1 (*EEF1A1*) gene.

1.4.1 AAVS1 Locus

The AAVS1 locus was first identified as a preferred site for AAV integration (Kotin et al. 1992). It spans approximately 4-Kb on chromosome 19 (position 19q13.42) and overlaps exon one of the *PPP1R12C* gene (also known as MBS85), which encodes protein phosphatase 1 regulatory subunit 12C (PPP1R12C). AAVS1 appears to be evolutionarily conserved in mammals, with homologues identified in both mouse and African green monkey (Dutheil et al. 2004). Investigation of the human locus

revealed a DNase I hypertensive site, which is suggestive of an open chromatin structure. In addition, an insulator sequence, capable of preventing the spread of heterochromatin, was identified at this locus (Ogata et al. 2003). More recently Sadelain and colleagues (2012) proposed AAVS1 as a potential 'safe harbour' locus for transgene integration. It has been shown to be transcriptionally active, lying within an open chromatin environment amenable to transgene expression in both iPSCs and haematopoietic stem cells (van Rensburg et al. 2013). While integration here disrupts *PPP1R12C*, no adverse consequences have been reported following integration. In their study, Henckaerts and colleagues (2009), demonstrated that AAV-mediated insertion at AAVS1 results in a partial duplication at the target site, thus preserving the *PPP1R12C* gene. Other studies using AAV-mediated transgene integration have also made use of AAVS1. Recchia et al. (2004) showed that transgenes can be successfully delivered by adenoviral-AAV hybrid vectors in a mouse model as a proof of concept for potential gene therapy.

AAVS1 has been extensively studied for its ability to support transgene integration and expression. Chang and Bouhassira (2012) demonstrated the ability of AAVS1 to support integration of a globin transgene following ZFN targeting in iPSCs, as a proof of concept for the treatment of α-thalassemia. Transgenes integrated here show robust, long-term expression; DeKelver and colleagues (2010) demonstrated that transgene constructs integrated at this locus exhibit prolonged expression for at least 50 generations in a range of cell types. In addition, Smith et al. (2008) were able to demonstrate expression of transgenes here even after stem cell differentiation.

More recently studies using TALENs have further demonstrated the use of this site for stable transgene expression in a range of cells. Cerbini at al. (2015) showed that AAVS1 could be used as a site for the efficient generation of modified iPSC cells. In addition, transgene expression at this site has been extensively characterised, particularly in human pluripotent stem cells (Tiyaboonchai et al. 2014).

1.4.2 hROSA26 Locus

The hROSA26 locus was first identified as the human homologue of the mouse *Rosa26* locus (Irion et al. 2007). It is located on human chromosome 3 (position 3p25.3) in close proximity to the *THUMPD3* gene, similar to the mouse *Rosa26*

which overlaps *ThumpD3* (Irion et al. 2007). The hROSA26 locus encodes a long non-coding RNA, THUMPD3-AS1, whose function is yet to be elucidated. Since its discovery over two decades ago in a retroviral gene-trapping screen (Friedrich & Soriano 1991), the mouse *Rosa26* locus has been commonly used in the creation of transgenic animals, as it demonstrates ubiquitous expression during development and is not prone to transgene silencing (Zambrowicz et al. 1997). Initial studies of the human ROSA26 locus have demonstrated its potential as a transgene integration site. Irion and colleagues (2007) showed that it has a broad expression pattern in adult tissues and stem cells; transgenes integrated here could be expressed in all three germ cell layers following stem cell differentiation. This locus is also conserved, with homologues identified in both rat (Kobayashi et al. 2012) and pig (Li et al. 2014), both of which show potential as transgene integration sites. Studies of the mouse *Rosa26* locus have revealed that expression is both promoter and orientation-dependent; some promoters demonstrate significant differences in expression in a sense compared to antisense orientation (Chen et al. 2011).

1.4.3 *IMPDH2* Gene

The *IMPDH2* locus also located on chromosome 3 (position 3p21.31) encodes inosine monophosphate dehydrogenase 2 (IMPDH2), responsible for catalysing the first committed and rate-limiting step of guanine nucleotide biosynthesis. It is believed to be essential for cell survival, and IMPDH2 null mice die in early embryogenesis (Gu et al. 2003). In humans there are two IMPDH isoforms, IMPDH1 and IMPDH2 which both carry out the same function and are expressed to varying degrees in most tissues. Both isoforms can be specifically inhibited by the drug mycophenolic acid (MPA), making this an attractive locus for the development of a drug-selectable transgene integration system (discussed in Chapter 4).

1.4.4 *EEF1A1* Gene

EEF1A1 encodes the eukaryotic translation elongation factor 1 alpha 1 (eEF1A1). It is located on chromosome 6 (position 6q14). Humans also have a second eEF1A isoform, eEF1A2, encoded by *EEF1A2* on chromosome 20. They are both responsible for providing aminoacyl-tRNA (aa-tRNA) molecules to the ribosome during the translation phase of protein synthesis. EEF1A1 is almost ubiquitously

expressed (Cho et al. 2012), except in a few tissues in which it is replaced by *EEF1A2* during early postnatal development. The eEF1A1 isoform is believed to be expressed in all immortalised cell lines, as such could provide a good site for transgene integration. The two proteins are 92% identical at the amino acid level, have very similar structures (Soares et al. 2009), and translation efficiencies (Kahns 1998). The reasons why eEF1A1 is replaced by eEF1A2 in a handful of tissues remain to be elucidated, but may potentially be due to differences in their moonlighting capacities. In addition to translation, eEF1A has variously been implicated in transformation susceptibility (Tatsuka et al. 1992), protein degradation and apoptosis (Cho et al. 2012), interactions with the cytoskeleton (Condeelis 1995), viral propagation (Mateyak & Kinzy 2010), and ageing (Shepherd et al. 1989). To what extent each isoform is involved remains to be confirmed.

1.5 Applications

My laboratory, as part of the UK cystic fibrosis (CF) gene therapy consortium (UKCFGTC), is interested in the development of recombinant lentiviral vectors for use in forthcoming CF clinical trials, which build upon previous trials using a lipid-based vector (Alton et al. 2015). As such, it would be advantageous to develop a lentiviral producer cell line, which could be used for the production of large amounts of clinical-grade lentiviral vectors for use in future clinical trials. Once loci, which can support robust long-term transgene expression, have been identified and validated, they could be used – along with validated SSNs – to express components of the lentiviral genome, which would provide a source of clinical grade lentiviral vectors.

Current methods of lentiviral vector production generally rely on transient transfection with a DNA cocktail encoding the components of the lentiviral genome, typically encoded on three or four separate plasmids, along with the transgene of interest. While this successfully generates high lentiviral titres, this method often shows poor reproducibility, scale up can be difficult and it requires large amounts of highly pure plasmid DNA, which can be challenging and costly to generate.

One of the problems of lentiviral production is that some viral components are cytotoxic – particularly the commonly used envelope protein, G protein from

vesicular stomatitis virus (VSV-G) – meaning that prolonged expression is impractical. To avoid this, many lentiviral vectors used to date in clinical trials have been derived from transient transfection (Aiuti et al. 2013; Biffi et al. 2013; Tebas et al. 2013; Cartier et al. 2009).

To overcome issues of cytotoxicity, inducible promoters can be used so that expression of cytotoxic viral components can be turned on only as required. Inducible stable producer lines have been successfully generated for some viral vectors (Greene et al. 2012; Throm et al. 2009a), however instability of the inducible promoters can lead to loss of viral production (Broussau et al. 2008; Ni et al. 2005; Farson et al. 2001). In addition, these methods require further purification steps to remove induction agents from the harvested virus.

Identification of loci, which can sustain prolonged transgene expression, and the development of tools (namely SSNs) with which to target these sites rapidly and precisely, will help to streamline future production of stable producer cell lines. SSN technologies will also aid the alteration of developed lines so that alternative envelope proteins or different transgenes can be utilised as desired.

Gene-specific editing can also be used in the study of gene function. In addition to their ability to drive the targeted insertion of transgenes, or other exogenous genetic material, SSNs can be used to selectively disrupt gene function. This occurs when no repair template is provided; the induced DSB is repaired by the cell's NHEJ repair pathway, which is error prone and often results in indels at the site of repair. These indels can lead to introduction of stop codons or frame shift mutations and ultimately loss of gene expression. This allows for the generation of knockout cell lines or transgenic animals in order to study the consequences of the loss of gene function. This can be achieved at both the single gene and genome wide level. All of the SSN platforms mentioned have been used to successfully generate specific gene knockouts in a range of cells and animal models (Carlson et al. 2012; Flemr & Bühler 2015; Gaj et al. 2012; Yang et al. 2014; Wettstein et al. 2015). However, the ability to multiplex the CRISPR/Cas9 system has allowed for the development of genome wide libraries, which can be used for recessive screens (Shalem et al. 2014a; Bassett et al. 2015; Koike-Yusa et al. 2013).

1.6 Aims

The main aim of this work was to develop SSNs – particularly TALENs and CRISPR/Cas9 – to target each of the four chosen loci and use these to investigate the potential of each site to support transgene expression and robust long-term expression, as a proof of principle for the development of stable lentiviral producer cell lines. In addition, I aimed to develop a drug-selectable transgene integration system at the *IMPDH2* locus, which would allow for rapid selection of correctly integrated transgenes. Finally, I aimed to use SSN technology to generate an *EEF1A1* knockout cell line, which I could use to investigate the potential moonlighting functions of the eEF1A2 isoform.

Chapter 2 Materials and Methods

2.1 General Methods

2.1.1 Polymerase Chain Reaction

DNA was amplified using the polymerase chain reaction (PCR). PCR was carried out using either Taq polymerase (Sigma-Aldrich®), Q5 high fidelity polymerase (NEB) or Herculase II Fusion polymerase (Agilent). The reaction compositions for each are given below:

Taq Polymerase	
Reaction buffer (10x)	1x
dNTP mix (10 mM)	200 µM (each)
Forward Primer (10 µM)	0.1 µM
Reverse Primer (10 µM)	0.1 µM
Taq	0.05 U
NF-Water	Up to 50 µL

Q5 High Fidelity Polymerase	
Reaction buffer (5x)	1x
dNTP mix (10 mM)	200 µM (each)
Forward Primer (10 µM)	0.5 µM
Reverse Primer (10 µM)	0.5 µM
Q5	0.02 U/µL
CG Enhancer (5x)	1x
NF-Water	Up to 50 µL

Herculase II Fusion Polymerase	
Reaction buffer (5x)	1x
dNTP mix (10 mM)	250 µM (each)
Forward Primer (10 µM)	0.25 µM
Reverse Primer (10 µM)	0.25 µM
Herculase	1 µL
NF-Water	Up to 50 µL

The cycling conditions used for each polymerase are listed below; annealing temperatures (Tm°) were determined based on primer composition.

Taq Polymerase

94°C	1 min	
94°C	1 min	30 cycles
Tm°	2 min	
72°C	90 sec	
72°C	5 min	
4°C	∞	

Q5 High Fidelity Polymerase

98°C	30 sec	
98°C	10 sec	30 cycles
Tm°	30 sec	
72°C	30 sec/Kb	
72°C	2 min	
4°C	∞	

Herculase II Fusion Polymerase

95°C	2 min	
95°C	20 sec	30 cycles
Tm°	20 sec	
72°C	30 sec/Kb	
72°C	3 min	
4°C	∞	

2.1.2 Manipulation of Nucleic Acids

2.1.2.1 Agarose Gel Electrophoresis

DNA was visualised using agarose gel electrophoresis. 1-2% agarose gels, according to the size of DNA fragments to be visualised, were made by dissolving UltraPure™ agarose (Invitrogen™) in 1x TBE buffer and heating. SYBR-Safe DNA Gel Stain (Invitrogen™) was added from a 10,000x stock to a final concentration of 1x. DNA samples were mixed with Orange G buffer prior to loading. DNA size markers were run alongside samples to estimate the size of DNA, routinely the 1-Kb Plus DNA ladder (Invitrogen™) and the HyperLadder™ 50bp (Bioline) were used. Electrophoresis was carried out at 100V in 1x TBE buffer.

2.1.2.2 DNA Purification

DNA was purified either using spin columns or by gel extraction methods. Following PCR, products were purified using the Wizard SV Gel and PCR Purification System Kit (Promega), according to the manufacturer's instructions. When DNA needed to

be separated from products of different sizes, gel extraction was carried out using the same kit, after excision of the band from the agarose gel. Purified DNA was eluted in NF-water.

2.1.2.3 DNA Quantification

DNA concentration was analysed using a NanoDrop1000 Spectrophotometer (Thermo Scientific). Purity was assessed using the ratio of absorbance at 260 nm and 280 nm wavelengths.

2.1.2.4 Digestion by Restriction Enzymes

All restriction enzymes were purchased from Roche or NEB. Reactions were carried out using optimal buffers, as recommended by the manufacturers. Double digests were only carried out where compatible buffers were available. All reactions were carried out in a 20 or 25 µL total volume and digestions were incubated at 37°C overnight.

2.1.2.5 Ligations

Fragments with cohesive ends were ligated overnight at 16°C using T4 DNA ligase (NEB). To maximise the ligation efficiency, 1:3 vector:insert molar ratios were used.

2.1.2.6 TOPO® Cloning

To sequence individual PCR amplicons, PCR products were cloned using the TOPO® TA for sequencing kit (Invitrogen™), according to the manufacturer's instructions. Briefly, 4 µL fresh PCR product (generated with a non-proofreading enzyme to ensure presence of 3'-deoxythymidine overhangs) was ligated into a pCR™4-TOPO TA vector, by incubation at room temperature for 5 minutes. Ligations were transformed using One Shot® TOP10 chemically competent *Escherichia coli* cells (Invitrogen™), according to the manufacturer's instructions.

2.1.2.7 Blunt-ended Cloning

Blunt-ended PCR products were cloned into the pJET1.2/blunt cloning vector, using a CloneJET PCR cloning kit (Thermo Scientific). An insert:vector ratio of 3:1 was used. Products were ligated using T4 DNA ligase, at room temperature for 30 minutes. Ligations were immediately transformed using DH5α chemically competent *E. coli* (Invitrogen™) as described in Section 2.3.1.

2.1.3 Sequencing

Sequencing of constructed TALENs was performed by DNA Sequencing & Services (MRC PPU, College of Life Sciences, University of Dundee, Scotland, www.dnaseq.co.uk) using Applied Biosystems Big-Dye version 3.1 chemistry on an Applied Biosystems model 3730 automated capillary DNA sequencer.

All other sequencing was carried out in house using Applied Biosystems Big-Dye version 3.1 chemistry. PCR products were first treated with Exo-SAP-IT (Affymetrix), 2 µL per 5 µL of PCR product, incubated at 37°C for 15 minutes followed by 80°C for 15 minutes.

The sequencing reaction was set up with the following components:

Template DNA	2 µL PCR product/ 200ng Plasmid
Big Dye	1 µL
Big Dye Buffer (5x)	1.5 µL
Primer (3.2 µM)	1.5 µL
NF-Water	up to 10 µL

These were incubated under the following cycling conditions:

96°C	1 min	
96°C	10 sec	30 cycles
50°	10 sec	
60°C	4 min	

Analysis was carried out by the sequencing technician Agnes Gallagher using an Applied Biosystems Genetic Analyzer, model 3130/3730. Interpretation of the results was carried out using Geneious® software (version 8.1.6).

2.1.4 Site-Directed Mutagenesis

Site directed mutagenesis was carried out to create the IMPDH^{IY} variant cDNA. Human *IMPDH2* cDNA was purchased from Source Bioscience (clone IRAK12 A06) and used as a template for site directed mutagenesis. This was carried out using the Q5 site directed mutagenesis kit (NEB). In addition, a codon optimised *T. foetus IMPDH* cDNA was synthesised and subcloned into the pcDNA3.1/Zeo(+) vector by GeneArt® (Life Technologies).

The following primers were used to introduce the sequence changes during the exponential amplification stage using Q5 polymerase:

Forward: 5′-CCAAGCAACAGCAGTGTACAAGGTGTACGAGTATGCACGGCGCTTT
Reverse: 5′-GGCCGCCCACAGGCCAGCACTTCCTGGATAATGCAGATGGAGCCAC

This was carried out using the following cycling conditions:

98°C	30 sec	
98°C	10 sec	25 cycles
63°	30 sec	
72°C	3 min	
72°C	2 min	
4°C	∞	

Amplified product was then subjected to kinase, ligase, DpnI (KLD) treatment; 1 µL PCR product was mixed with 5 µL 2x KLD reaction buffer and 1 µL KLD enzyme mix in a 10 µL reaction. This was incubated at room temperature for 5 minutes, then transformed into NEB 5-alpha competent *E. coli*, according to the manufacturer's instructions. Plasmids were verified by sequencing and re-isolated using the peqGOLD XChange Plasmid Maxi EF Kit (peqlab), to produce sufficient yield for subsequent applications.

2.1.5 Quantitative PCR

2.1.5.1 cDNA Synthesis

Synthesis of cDNA was carried out using a nanoScript 2 reverse transcription kit (Primerdesign) with an oligo dT primer, according to the manufacturer's instructions. For each RNA sample, a control with no reverse transcriptase was included, to assess potential genomic DNA contamination.

Following synthesis, a control PCR was carried out to check for potential contamination using the following primers:

Forward: 5′-CGGTGTCTGTAGTGGCTTGA
Reverse: 5′-TGTTTTTGACGACGAAGCAG

This resulted in a product of 238-bp from the cDNA template and 981-bp from a genomic DNA template. The PCR was carried out using the following cycling conditions:

98°C	30 sec	
98°C	10 sec	10 cycles
65°C (-1°C/cycle)	30 sec	
72°C	30 sec	
98°C	10 sec	30 cycles
55°C	30 sec	
72 °C	30 sec	
72 °C	2 min	
4 °C	∞	

Products were analysed using agarose gel electrophoresis. Any samples that contained genomic DNA contamination were discarded.

2.1.5.2 geNORM

A geNORM reference gene kit (Primerdesign) was used to identify appropriate reference genes. Reactions were carried out using a panel of six reference genes in all cDNA samples. The stability of each reference gene across all samples was calculated using qbase+ software (Biogazelle).

2.1.5.3 Quantitative PCR

Reactions were carried out using PrecisionPLUS master mix (Primerdesign) with SYBR green chemistry. 10 µL reaction volumes were used, with 2.5 µL cDNA at 1:200 dilution. Each sample was run in triplicate. Primers are listed in Table 2.1. Reactions were carried out on an Applied Biosystems 7900HT Fast Real-Time PCR detection system using the following cycling conditions:

95°C	2 min	
95°C	20 sec	40 cycles
60°C*	20 sec	
Melt curve		

* Data acquisition

Data analysis was carried out using SDS 2.3 software. Samples were excluded from analysis if there was a difference between their Ct value and the Ct values of the

other replicates in their group (each sample was run in triplicate) of greater than 1 Ct. Expression was calculated using the following formula:

Quantity = 10^((Ct value – Y-intercept)/Slope)

Y-intercept and Slope were determined from the standard curve for each primer pair.

Mean expression of each triplicate was normalised to the geometric mean of five reference genes, determined by geNORM analysis.

Table 2.1 qPCR primers. The AAVS1 and hROSA26 loci are not genes, therefore genes in these regions were selected to act as surrogate markers of transcriptional activity at these sites, given in parentheses. * Primers were designed by Primerdesign;† primer sequence taken from Irion et al. (2007).

Target Site	Primer	Sequence
AAVS1 (*PPP1R12C*)	Forward	CGCGAGAAGATTTCCCTCCA
	Reverse	CTCTTCAAGCTGCACGGGAC
hROSA26† (*THUMPD3-AS1*)	Forward	TTATCCGTTGCGTAAGCACAGAGAGG
	Reverse	AGCCTCTGCACACCGTGAGAATAA
IMPDH2	Forward	CCCAGGCCAAGAACCTCATT
	Reverse	GCCAGCATTCCTGCGTAA
EEF1A1*	Forward	TTTGGGTCGCTTTGCTGTTC
	Reverse	CTTCTTGTCCACTGCTTTGATGA
EEF1A2*	Forward	GTATTGACAAAAGGACCATTGAGAAG
	Reverse	CAGCACCCAGGCATACTTGA

2.1.6 Western Blotting

2.1.6.1 Protein Quantification and Normalisation

A BCA protein assay kit (Thermo Scientific) was used to measure the protein concentration of each sample. 10 µL of each sample was mixed with 200 µL working reagent (reagent A and reagent B, 50:1 ratio), in triplicate, in a 96-well plate on a plate shaker for 30 seconds, then incubated at 37°C for 30 minutes. Once cooled to room temperature, absorbance was measured at a wavelength of 562 nm using a plate reader. Protein concentration was calculated using a standard curve generated from a dilution series of protein standards. The concentration of all samples was normalised to that of the least concentrated sample by dilution with RIPA buffer. An equal volume of 2x Laemmli buffer was added to each sample, which were vortexed and incubated at 100°C for 7 minutes. 10% DTT was added from a 1 M stock solution and samples were stored at -20°C.

2.1.6.2 SDS-PAGE

The BioRad Mini PROTEAN 3 mini-gel apparatus was used according to the manufacturer's instructions. 15 µL of prepared protein samples were run on 10% SDS-PAGE gels (4.3% stacking gel) at 150 V at 4°C in 1x Tris/Glycine/SDS buffer (BioRad) until the dye front reached the bottom of the gel. The prestained broad range protein marker (NEB) was routinely run alongside experimental samples to estimate protein size.

2.1.6.3 Blotting

The BioRad Mini Trans-blot Electrophoretic Transfer Cell was used to carry out Western blotting. Immobilon®-FL PVDF membrane (Merck Millipore) was first pre-wet in 100% methanol. The membrane, gel, filter papers and sponges were all soaked in 1x transfer buffer. The sponge, filter paper, gel and membrane were all stacked and assembled into the apparatus. The tank was the filled with 1x transfer buffer and run at 100 V at 4°C for 1 hour. Membranes were dried at 40°C for 15 minutes.

2.1.6.4 SYPRO® Ruby Total Protein Staining

To ensure even and complete protein transfer, membranes were treated with SYPRO® Ruby Protein Gel stain (Thermo Scientific). All incubation steps were carried out on an orbital shaker at room temperature. Dried membranes were re-wet in 100% methanol. The membranes were then incubated in 7% (v/v) acetic acid/ 10% (v/v) methanol for 15 minutes, then washed four times in distilled water, 5 minutes per wash. The membranes were then incubated with SYPRO® Ruby for 15 minutes in the dark and then washed four times with distilled water for 1 minute each. The membranes were then visualised using the Odyssey® Fc imaging system (LI-Cor) at a wavelength of 600 nm.

2.1.6.5 Antibody Probing of Western Membranes

All incubation steps were carried out on an orbital shaker at room temperature, unless otherwise stated. Membranes were blocked by incubation in Odyssey® Blocking Buffer (PBS) (LI-Cor) mixed 1:1 with PBS, at 4°C overnight. Membranes were then incubated with primary antibodies diluted in Odyssey® Blocking Buffer (PBS) for 2 hours. Membranes were then washed four times in PBS-T, 5 minutes per

wash. Membranes were then incubated with fluorescent secondary antibodies (LI-Cor), diluted 1:5000 in Odyssey® Blocking Buffer (PBS), for 1 hour. Finally, the membranes were washed four times in PSB-T, 5 minutes per wash and visualised using the Odyssey® Fc imaging system, at 600 nm, 700 nm, and 800 nm wavelengths as appropriate.

Image analysis and quantification was carried out using Image Studio Lite (version 4.0); for each sample band intensities were normalised to the band intensities of the loading control for each sample (GAPDH or Tubulin).

Table 2.2 Western Blot antibodies.

Antibody	Working Dilution
Sheep anti-eEF1A1	1:2000
Mouse anti-Myc (Santa Cruz)	1:2000
Mouse anti-GAPDH (Merck Millipore)	1:1000
Rabbit anti-Tubulin (ABCAM)	1:5000
Anti-rabbit (LI-Cor)	1:5000
Anti-mouse (LI-Cor)	1:5000
Anti-goat (LI-Cor)	1:5000

2.2 SSN Construction and Validation

2.2.1 TALEN Construction

Target sites were identified using the online software available at https://tale-nt.cac.cornell.edu/node/add/talen. All TALENs were constructed using the protocol developed by the Zhang laboratory (Sanjana et al. 2012), which is outlined below. All plasmids used in TALEN construction (Table 2.3) were a gift from Feng Zhang (Addgene, Kit #1000000019).

Table 2.3 Plasmids used in TALEN Construction

Plasmid	Description
pNI_v2	Monomer plasmid with NI RVD sequence
pNG_v2	Monomer plasmid with NG RVD sequence
pNN_v2	Monomer plasmid with NN RVD sequence
pHD_v2	Monomer plasmid with HD RVD sequence
pTALEN_v2 (NI)	Backbone plasmid with NI 0.5 repeat
pTALEN_v2 (NG)	Backbone plasmid with NG 0.5 repeat
pTALEN_v2 (NN)	Backbone plasmid with NN 0.5 repeat
pTALEN_v2 (HD)	Backbone plasmid with HD 0.5 repeat

2.2.1.1 Monomer Library Construction

Each monomer was amplified from a plasmid template (Table 2.3) using primers encoding ligation adapters, which define its position in the final monomer array (Table 2.4). Each monomer was amplified with 18 primer pairs, such that it could be positioned anywhere in the final array. An additional amplification was carried out to allow for construction of 17-mer arrays, required in the construction of the AAVS1-targeting TALEN pair. This resulted in a library of 76 monomers. The primer mixes required are listed in Table 2.4.

Table 2.4 Monomer library construction primers. Each primer pair adds adapters to the monomers such that their position in the final array is specified. For a 17-mer array, the final primer pair is altered.

Forward Primer	Reverse Primer	Position in Array
Ex-F1	In-R1	1
In-F2	In-R2	2
In-F3	In-R3	3
In-F4	In-R4	4
In-F5	In-R5	5
In-F6	Ex-R1	6
Ex-F2	In-R1	7
In-F2	In-R2	8
In-F3	In-R3	9
In-F4	In-R4	10
In-F5	In-R5	11
In-F6	Ex-R2	12
Ex-F3	In-R1	13
In-F2	In-R2	14
In-F3	In-R3	15
In-F4	In-R4	16
In-F5	In-R5	17
In-F5	Ex-R4	17 (17-mer array)
In-F6	Ex-R3	18

Monomers were amplified using the following components:

Monomer template plasmid (5 ng/µL)	1 µL
dNTPs (10 mM)	1 µL
Herculase II PCR buffer (5x)	20 µL
Primer mix (10 µM each)	2 µL
Herculase II Fusion polymerase	1 µL
NF-Water	75 µL

The reaction was performed using the following conditions:

95°C	2 min	
95°C	20 sec	40 cycles
60°C	20 sec	
72°C	10 sec	
72°C	3 min	
4°C	∞	

Each reaction was run in duplicate to give a greater yield. Amplified monomers were analysed by agarose gel electrophoresis, to ensure correct amplification. Duplicate samples were pooled and purified using the QIAquick® PCR purification kit (Qiagen®) according to the manufacturer's instructions. Samples were eluted in 100 µL Buffer EB. Each monomer was diluted to give a final concentration of 15 ng/µL (monomers 1, 6, 7, 12, 18 and 18 were diluted to 18 ng/µL as these are slightly longer than the other monomers). The monomer library was stored at -20°C.

2.2.1.2 Construction of Custom TALENs

Target sequences were divided into three hexamers (for 18-mer arrays) or two hexamers and a pentamer (for 17-mer arrays). Each hexamer was assemble by a Golden Gate reaction, with the following components:

BsmBI (10 U/µL)	0.75 µL
Tango buffer (10x)	1 µL
DTT (10 mM)	1 µL
T4 DNA ligase (2000 U/µL)	0.25 µL
ATP (10 mM)	1 µL
Six monomers	1 µL each

Reactions were incubated at 37°C for 5 minutes, then 20°C for 5 minutes for 20 cycles. Ligations were verified by agarose gel electrophoresis. To remove unsuccessfully ligated products, the hexamers were treated with PlasmidSafe exonuclease, in the presence of ATP (1 mM final concentration) for 30 minutes at 37°C followed by inactivation at 70°C for a further 30 minutes.

Next, each hexamer was PCR amplified using Herculase II Fusion polymerase (Section 2.1.1), using Hex-F and Hex-R primers (Table 2.5). Products were electrophoresed using a 2% agarose gel, bands of the correct size were excised and DNA purified using the MiniElute gel extraction kit (Qiagen®) according to the

manufacturer's instructions. Products were eluted in 20 µL Buffer EB. Concentration was determined using a NanoDrop1000 Spectrophotometer and the concentration of each hexamer (or pentamer) was normalised by dilution with Buffer EB to a final concentration of 20 ng/µL.

The hexamers were assembled into the appropriate backbone plasmid using a second Golden Gate reaction using the following components:

TALE Backbone vector (100 µng/L)	1 µL
BsaI-HF (20 /µL)	0.75 µL
NEBuffer 4 (10x)	1 µL
BSA (10x)	1 µL
ATP (10 mM)	1 µL
T4 DNA ligase (2000 U/µL)	0.25 µL
Three purified hexamers (20 ng/µL each)	1 µL each
NF-Water	2 µL

The reaction was performed using the following cycling conditions:

37°C	5 min	20 cycles
20°C	5 min	
80°C	20 min	

The constructed TALENs (5 µL Golden Gate reaction) were then transformed into 50 µL DH5α chemically competent *E. coli* (Section 2.3.1). A colony PCR was carried out to identify colonies containing correctly constructed TALENs. Individual colonies were picked and streaked on to LB-amp plates, then swirled into 100 µL NF-Water. PCR was carried out using Taq (Sigma-Aldrich®) with TALE-Seq-F1 and TALE-Seq-R1 primers (Table 2.5) on 1 µL of colony suspension. The following cycling conditions were used:

94°C	3 min	
94°C	30 sec	30 cycles
60°C	30 sec	
68°C	2 min	
68°C	5 min	
4°C	∞	

Products were examined following agarose gel electrophoresis. Positive colonies were inoculated into 5 mL LB with ampicillin and grown overnight at 37°C,

225 rpm. Plasmids were isolated using the QIAprep Spin miniprep kit (Qiagen®) according to the manufacturer's instructions. Correct assembly of TALENs were confirmed by sequencing.

Table 2.5 Primers for TALEN Construction

Forward Primer	Sequence
Ex-F1	5′-TGCGTCcgtctcCGAACCTTAAACCGGCCAACATACCggtctcCTGACCCCAGAGCAGGTCGTG
Ex-F2	5′-TGCGTCcgtctcCGAACCTTAAACCGGCCAACATACCggtctcGACTTACACCCGAACAAGTCGTGGCAATTGCGAGC
Ex-F3	5′-TGCGTCcgtctcCGAACCTTAAACCGGCCAACATACCggtctcGCGGCCTCACCCCAGAGCAGGTCG
Ex-R1	5′-GCTGACcgtctcCGTTCAGTCTGTCTTTCCCCTTTCCggtctcTAAGTCCGTGCGCTTGGCAC
Ex-R2	5′-GCTGACcgtctcCGTTCAGTCTGTCTTTCCCCTTTCCggtctcAGCCGTGCGCTTGGCACAG
Ex-R3	5′-GCTGACcgtctcCGTTCAGTCTGTCTTTCCCCTTTCCggtctcTCCCATGGGCCTGACATAACACAGGCAGCAACCTCTG
Ex-R4	5′-GCTGACcgtctcCGTTCAGTCTGTCTTTCCCCTTTCCggtctcTGAGTCCGTGCGCTTGGCAC
In-F2	5′-CTTGTTATGGACGAGTTGCCcgtctcGTACGCCAGAGCAGGTCGTGGC
In-F3	5′-CCAAAGATTCAACCGTCCTGcgtctcGAACCCCAGAGCAGGTCGTG
In-F4	5′-TATTCATGCTTGGACGGACTcgtctcGGTTGACCCCAGAGCAGGTCGTG
In-F5	5′-GTCCTAGTGAGGAATACCGGcgtctcGCCTGACCCCAGAGCAGGTCGTG
In-F6	5′-TTCCTTGATACCGTAGCTCGcgtctcGGACACCAGAGCAGGTCGTGGC
In-R1	5′-TCTTATCGGTGCTTCGTTCTcgtctcCCGTAAGTCCGTGCGCTTGGCAC
In-R2	5′-TCTTATCGGTGCTTCGTTCTcgtctcCCGTAAGTCCGTGCGCTTGGCAC
In-R3	5′-TGAGCCTTATGATTTCCCGTcgtctcTCAACCCGTGCGCTTGGCACAG
In-R4	5′-AGTCTGTCTTTCCCCTTTCCcgtctcTCAGGCCGTGCGCTTGGCACAG
In-R5	5′-CCGAAGAATCGCAGATCCTAcgtctcTTGTCAGTCCGTGCGCTTGGCAC
Hex-F	5′-CTTAAACCGGCCAACATACC
Hex-R	5′-AGTCTGTCTTTCCCCTTTCC
TALE-Seq-F1	5′-CCAGTTGCTGAAGATCGCGAAGC
TALE-Seq-F2	5′-ACTTACACCCGAACAAGTCG
TALE-Seq-R1	5′-TGCCACTCGATGTGATGTCCTC
TALE-Seq-R2	5′-CCCATGGGCCTGACATAA

2.2.2 Construction of sgRNAs

All sgRNAs were constructed using the protocol developed by the Zhang laboratory (Ran, P D Hsu, et al. 2013). All plasmids used in sgRNA construction were a gift

from Feng Zhang (Addgene plasmids #48139 and #48141[1]). Target sites were identified using online software available at http://crispr.mit.edu/ (Cong et al. 2013). Oligonucleotides encoding the top and bottom strands of the 20-bp target with ligation adapters were ordered from Sigma-Aldrich:

Top: 5′-CACCGNNNNNNNNNNNNNNNNNNN
Bottom: 5′-AAACNNNNNNNNNNNNNNNNNNC

They were annealed and phosphorylated using T4 PNK at 37°C for 30 minutes, followed by 95°C for 5 minutes then cooled to 25°C at -5°C per minute.

Annealed and phosphorylated oligonucleotides were diluted 1:200 in NF-water, then cloned into a backbone plasmid (either pX459 or pX462) using a Golden Gate reaction with the following components:

Backbone plasmid (100 ng/µL)	1 µL
Diluted oligo duplex	2 µL
Tango buffer (10x)	2 µL
DTT (10 mM)	1 µL
ATP (10 mM)	1 µL
FastDigest BbsI	1 µL
T4 Ligase	0.5 µL
NF-Water	11.5 µL

This was incubated at 37°C for 5 minutes then 21°C for 5 minutes for six cycles.

Ligated products were treated with PlasmidSafe exonuclease, in the presence of ATP (1 mM final concentration) for 30 minutes at 37°C, followed by inactivation at 70°C for a further 30 minutes.

Constructed plasmids were then transformed into DH5α chemically competent *E. coli* (Section 2.3.1). Plasmids were isolated using the peqGOLD Plasmid Miniprep kit (C-line) (peqlab), according to the manufacturer's instructions. Correct construction was confirmed by sequencing using the U6-Fwd primer:

5′-GAGGGCCTATTTCCCATGATTCC

[1] Since the completion of this work, a point mutation in the puromycin resistance gene has been identified in these plasmids, which makes it less effective in some cell lines. Newer versions are now available (Addgene plasmids #62988 and #62987).

2.2.3 TALEN *in vitro* Cleavage Assay

TALEN proteins were produced *in vitro* using the T_NT® Quick Coupled Transcription/Translation System (Promega), according to the manufacturer's instructions. To visualise proteins, reactions were carried out with and without Transcend™ Biotin-Lysyl-tRNA, so that the resulting proteins contained biotinylated lysine residues. Correct production was confirmed by western blot. Reactions were mixed with 2x Laemmli buffer, separated by SDS-PAGE, and transferred to membranes (section 2.1.6.3). Membranes were incubated in Western Blue® Stabilized Substrate for Alkaline Phosphatase, to visualise biotinylated proteins.

Target DNA was amplified by PCR using Q5 polymerase, using the following primers:

AAVS1_Fw: 5′-ACACCTAGGACGCACCATTC
AAVS1_Rv: 5′-CTTGCTTTCTTTGCCTGGAC
hROSA26_Fw: 5′-GCAAGGGTAGCATTCCAAAA
hROSA26_Rv: 5′-CTTGGGTAGCCTGGTGTCTC

The *in vitro* cleavage assay was carried out with non-biotinylated TALENS, using the following components:

DNA target	200 ng
BSA (10x)	1x
NEBuffer 4 (10x)	1x
NaCl (1 M)	0.1 M
Right TALEN	1 µL
Left TALEN	1 µL
NF-Water	up to 10 µL

These were incubated at 37°C for 90 minutes. Reactions were analysed using agarose gel electrophoresis.

2.2.4 Surveyor Assay

All assays were carried out using the SURVEYOR mutation detection kit for standard gel electrophoresis (Transgenomic).

2.2.4.1 Assay Optimisation

For optimisation, the kit's internal controls were used. PCR amplicons were generated from two plasmid templates – Control C and Control G – using the following primers:

Forward: 5′-ACACCTGATCAAGCCTGTTCATTTGATTAC
Reverse: 5′-CGCCAAAGAATGATCTGCGGAGCTT

PCR was carried out using either Taq polymerase or Herculase II Fusion polymerase (Section 2.1.1), using an annealing temperature of 65°C. This resulted in two amplicons, which differed by a single base. In addition, amplification was carried out using Herculase II Fusion polymerase using PCR buffer from Sigma-Aldrich®. PCR products were analysed by agarose gel electrophoresis, to ensure correct size. PCR products were purified using the Wizard SV Gel and PCR Purification System Kit, according to the manufacturer's instructions, and were eluted in Surveyor buffer. The concentration was determined using a NanoDrop 1000 Spectrophotometer and were diluted to 50 ng/µL in Surveyor Buffer.

Samples were supplemented with 0.5x, 1x or 2x the standard concentration of Herculase II Fusion polymerase before or after cross-hybridisation. Cross-hybridisation was carried out in a thermal cycler under the following conditions:

95°C, 10 minutes
95-85°C, -2°C/second
85°C, 1 minute
85-75°C, -0.3°C/second
75°C, 1 minute
75-65°C, -0.3°C/second
65°C, 1 minute
65-55°C, -0.3°C/second
55°C, 1 minute
55-45°C, -0.3°C/second
45°C, 1 minute
45-35°C, -0.3°C/second
35°C, 1 minute
35-25°C, -0.3°C/second
25°C, 1 minute
4°C, ∞

Following cross hybridisation, DNA was digested with Surveyor nuclease, using the following components:

$MgCl_2$ (0.15 M)	2 µL
Surveyor nuclease S	1µL
Surveyor enhancer S	1 µL
Reannealed duplex DNA	4 µL
Surveyor buffer	12 µL

The reaction was incubated at 42°C for 1 hour, and then 2 µL Stop solution was added. 10 µL of each reaction was electrophoresed on a 2% agarose gel. Gels were visualised using the Odyssey® Fc imaging system at a wavelength of 600 nm. Band intensities were calculated using Image Studio Lite. The cleavage fraction (f_{cut}) was calculated using the following equation:

$f_{cut} = a / (a + b)$

a = the integrated intensity of both of the cleavage product bands and b = the integrated intensity of the uncleaved PCR product band.

2.2.4.2 Optimised Protocol

Target sites were amplified by PCR using Taq polymerase (Section 2.1.1). Primers and annealing temperatures are listed in Table 2.6.

Table 2.6 Target-specific Primers

Name	Sequence	Tm°
AAVS1_Fw	5′-TCTGGCAAGGAGAGAGATGG	60°C
AAVS1_Rv	5′-CTCTAGTCTGTGCTAGCTCTTCCAG	
hROSA26_Fw	5′-GTCGCATTCAAAGCTGTCCT	60°C
hROSA26_Rv	5′-CAGAGAATTGGATCTGAGGTTAGG	
IMPDH2_Fw	5′-GGGAAAGGAGTGTCAAACCA	60°C
IMPDH2_Rv	5′-TACGTGGGAGGTGAGATGTG	
EEF1A1_Ex3_Fw	5′-GGAGGCTGCTGAGGTATGTT	59.5°C
EEF1A1_Ex3_Rv	5′-CCAACCTGAGATGTCCCTGT	
EEF1A1_Ex4_Fw	5′-AAGGCTGACTGTGCTGTCCT	59.5°C
EEF1A1_Ex4_Rv	5′-CAGCCAACTTACCACCAATTT	

Products were analysed using agarose gel electrophoresis, and diluted to approximately equal concentrations using 1x PCR buffer.

Cross-hybridisation was performed using the conditions detailed in Section 2.2.4.1. Reannealed duplexes were then digested with Surveyor nuclease, using the following components:

$MgCl_2$ (0.15 M)	2 µL
Surveyor nuclease S	1µL
Surveyor enhancer S	1 µL
Reannealed duplex DNA	16 µL

The reaction was incubated at 42°C for 1 hour, then 2 µL stop solution was added. 10 µL of each reaction was electrophoresed on a 2% agarose gel. Gels were visualised using the Odyssey® Fc imaging system at a wavelength of 600 nm. Band intensities were calculated using Image Studio Lite. The cleavage fraction (f_{cut}) was calculated as before. This was used to estimate the percentage of SSN-mediated modification using the following equation:

$100 \times (1 - (1 - f_{cut})^{1/2})$

2.2.5 Targeting Cassette Construction

2.2.5.1 AAVS1 Cassette

The AAVS1 transgene integration cassette was constructed using long multiple fragment PCR based on the method described by Shevchuk et al. (2004). The homology arms were amplified from HEK293FT genomic DNA. The *EGFP* fragment was amplified from the pEGFP_N1 plasmid.

Table 2.7 Primers with adapters used to amplify the three fragments. Adapter sequences are given in lowercase. Primers were made up to 50 µM in 5mM Tris-HCl pH 8.5.

Name	Sequence
EGFP_Fw	5′-ccaatatcaggagactaggaaggaggaggcctaaggatggggctttt tctgtcaccaatcctgtccctagtCTTGAGCGTCGATTTTTGTGATGCTC
EGFP_Rv	5′-tcccagggccggttaatgtggctctggttctgggtacttttatctg tcccctccaccccacagtggggccACGGCACCTCGACCCCAAAA
AAVS1_Left_Fw	5′-AGAGCAGAGCCAGGAACCCCTGTAG
AAVS1_Left_Rv	5′-ACTAGGGACAGGATTGGTGACAGAAAAGC
AAVS1_Right_Fw	5′-GGCCCCACTGTGGGGTGG
AAVS1_Right_Rv	5′-CTCTCCTGAGTCCGGACCACTTTGA

The Left arm was amplified using the following components:

dNTPs (10 mM)	2.5 µL
Herculase II Fusion Buffer (5x)	10 µL
gDNA (10ng/ µL)	5 µL
AAVS1_Left_Fw Primer (50 µM)	1 µL
AAVS1_Left_Rv Primer (50 µM)	1 µL
Herculase II Fusion Polymerase	1 µL
DMSO	1.5 µL
NF-Water	28 µL

The reaction was performed using the following cycling conditions:

95°C	2 min	
95°C	10 sec	26 cycles
68°C	30 sec	
72°C	1 min	
72°C	3 min	
4°C	∞	

The right arm was amplified using the following components:

dNTPs (10 mM)	2.5 µL
Herculase II Fusion Buffer (5x)	10 µL
gDNA (10ng/ µL)	5 µL
AAVS1_Right_Fw Primer (50 µM)	1 µL
AAVS1_Right_Rv Primer (50 µM)	1 µL
Herculase II Fusion Polymerase	1 µL
DMSO	1.5 µL
NF-Water	28 µL

This was performed using the same cycling conditions as the left arm, but with an annealing temperature of 69°C.

The *EGFP* fragment was amplified using the following components:

dNTPs (10 mM)	2.5 µL
Herculase II Fusion Buffer (5x)	10 µL
pEGFP_N1 (1 µg/mL)	0.3 µL
EGFP_Fw (50 µM)	1 µL
EGFP_Rv (50 µM)	1 µL
Herculase II Fusion Polymerase	1 µL
DMSO	1.5 µL
NF-Water	32.7 µL

The reaction was performed using the following conditions:

95°C	2 min	
95°C	10 sec	27 cycles
64°C	30 sec	
72°C	2 min 30 sec	
72°C	3 min	
4°C	∞	

Correct amplification was confirmed using agarose gel electrophoresis and the products were purified using QIAquick® PCR purification kit, according to the manufacturer's instructions, and eluted in 30 µL of 5mM Tris-HCl, pH 8.5.

The three fragments were fused using the following components:

NF-Water	8.25 µL
Reaction buffer (5x)	10 µL
dNTP mix (10 mM)	5 µL
Q5	0.5 µL
CG Enhancer (5x)	1 µL
Left arm	80 ng
Right arm	90 ng
EGFP fragment	200 ng

This was incubated using the following cycling conditions:

98°C	30 sec	
98°C	10 sec	15 cycles
56°C	30 sec	
72°C	1 min 30 sec	
72°C	2 min	
4°C	∞	

Correct fusion was confirmed by agarose gel electrophoresis. The product was amplified using the following nested primers:

Nested_AAVS1_Fw: 5′-CCAAGGACTCAAACCCAGAA
Nested_AAVS1_Rv: 5′-TTGCTCTCTGCTGTGTTGCT

This was carried out using Q5 polymerase, with an annealing temperature of 60°C. The final product was confirmed using agarose gel electrophoresis and cloned into pJET1.2/blunt vector using the CloneJET PCR cloning kit, according to the manufacturer's instruction. This was then transformed into DH5α *E. coli*. Plasmids

were isolated and correct assembly was confirmed by sequencing (primers listed in Table 2.10).

2.2.5.2 IMPDH2 Cassette

Homology arms were amplified from HEK293FT genomic DNA with Q5 polymerase using the primers listed in Table 2.8.

Table 2.8 IMPDH2 homology arm primers. Restriction enzyme adapters are indicated in lowercase.

Name	Sequence	Tm°
IMPDH2_Left_Fw	5′-ggggctcgagGACTCACAGCACAGCAGCTC	62°C
IMPDH2_Left_Rv	5′-gggggcggccgcGGATGCAGGAACCCAATAAA	
IMPDH2_Right_Fw	5′-gggggaattcGACTGAGGCAGGCAAATCAT	66°C
IMPDH2_Right_Rv	5′-gggggtaccGACCTCTACTCACCCCCACA	

The right homology arm and pSicoR-Ef1a-mCh-Puro vector (a gift from Bruce Conklin; Addgene plasmid #31845) were digested with EcoRI and KpnI enzymes and purified by gel extraction using the Wizard SV Gel and PCR Purification System Kit, according to the manufacturer's instructions. These were ligated using T4 DNA ligase and then transformed into DH5α *E. coli*. The resulting plasmid (pSicoR-Ef1a-mCh-Puro-IMPDH2_Right) was isolated and correct ligation was confirmed using a diagnostic digest.

The left arm and the pSicoR-Ef1a-mCh-Puro-IMPDH2_Right plasmid were digested with XhoI and NotI, then gel purified as before. These were ligated using T4 DNA ligase and transformed into DH5α *E. coli.* Following plasmid isolation, correct ligation was confirmed by sequencing (primers listed in Table 2.10).

2.2.5.3 hROSA26 Cassette

Homology arms were amplified from HEK293FT genomic DNA with Q5 polymerase using the primers listed in Table 2.9.

Table 2.9 hROSA26 homology arm primers. Restriction enzyme adapters are indicated in lowercase.

Name	Sequence	Tm°
hROSA26_Left_Fw	5′-ggggctcgagAAAGGGAGATGTTGCCTAACC	66°C
hROSA26_Left_Rv	5′-gggggcggccgcgctagcGCAATCCTGAGGGAGCTG	
hROSA26_Right_Fw	5′-ggggcaattgTTTACATGTCGCTCACTCCA	59°C
hROSA26_Right_Rv	5′-gggggtaccGAGCATGAAAATGACAGGTGAA	

The left homology arm and pSicoR-Ef1a-mCh-Puro vector were digested with XhoI and NotI and purified by gel extraction using the Wizard SV Gel and PCR Purification System Kit, according to the manufacturer's instructions. These were ligated using T4 DNA ligase and then transformed into DH5α *E. coli*. The resulting plasmid (pSicoR-Ef1a-mCh-Puro-hROSA26_Left) was isolated and correct ligation was confirmed using a diagnostic digest.

The right arm was digested with MfeI and KpnI and the pSicoR-Ef1a-mCh-Puro-hROSA26_Left plasmid was digested with EcoRI and KpnI, then gel purified as before. These were ligated using T4 DNA ligase and transformed into DH5α *E. coli.* Following plasmid isolation, correct ligation was confirmed by sequencing (primers listed in Table 2.10).

Table 2.10 Sequencing primers for targeting cassettes

Primer	Sequence
AAVS1_Seq_F1	5′-TTACGGTTCCTGGCCTTTTG
AAVS1_Seq_F2	5′-ATAGCGGTTTGACTCACGGG
AAVS1_Seq_F3	5′-CTACCCCGACCACATGAAGC
AAVS1_Seq_R1	5′-CAAAGTACCCCGTCTCCCTG
AAVS1_Seq_R2	5′-ACAACACTCAACCCTATCCGG
hROSA26_Seq_R1	5′-TTCTGTCCCTGACCGTCTTC
hROSA26_Seq_R2	5′-TCTTGCTCAACTGGTACTAGCT
hROSA26_Seq_R3	5′-TCATTCACCTGGGGCTGATT
hROSA26_Seq_R4	5′-CTTCTCTAGGCACCGGTTCA
hROSA26_Seq_R5	5′-CATGTTATCCTCCTCGCCCT
hROSA26_Seq_R6	5′-CTCTGCTTGATCTCGCCCTT
hROSA26_Seq_R7	5′-TGGGGTTGTTGTCATTGCTG
hROSA26_Seq_R8	5′-GCACCATCCAAAGGTCAGTG
IMPDH2_Seq_F1	5′-CTTCTGGGGTGCTGATAGGG
IMPDH2_Seq_F2	5′-CCTGGTCTCTGGGCTAACTC
IMPDH2_Seq_F3	5′-TCCTGGGAATTGCACTGAGG
IMPDH2_Seq_F4	5′-TCTAGCAGCCCAGGTTGAAG
IMPDH2_Seq_F5	5′-TGCCTCCTGAACTGCGTC
IMPDH2_Seq_F6	5′-GCGTGATGAACTTCGAGGAC
IMPDH2_Seq_F7	5′-TCGAACCGAGTACAAGCCC
IMPDH2_Seq_F8	5′-CTCCCCTTCTACGAGCGG
IMPDH2_Seq_F9	5′-CATCATTGCCCGGACAGAC
IMPDH2_Seq_F10	5′-ATGGGAAGTGGCTCCATCTG
IMPDH2_Seq_R1	5′-GGGAAGGGAGTGAATTGGGA

2.2.5.4 Integration Detection

Correct integration of targeting cassettes was determined by PCR using Q5 polymerase. The primers used are listed in Table 2.11.

Table 2.11 Primers used to detect correct integration of targeting cassettes. The pSicoR_Rv primer was used for both hROSA26 and IMPDH2.

Primer	Sequence
hROSA26_Fw	5′-TGGTTTGTGGGTGGGAGG
IMPDH2_Fw	5′-CCATGGCCGACTACCTGATT
pSicoR_Rv	5′-CTTCTCTAGGCACCGGTTCA
AAVS1_Integration_Fw	5′-ACACCTAGGACGCACCATTC
AAVS1_Integration_Rv	5′-AGGTTCCGTCTTCCTCCACT

2.3 Bacterial Methods

2.3.1 Plasmid Transformation

Chemically competent sub-cloning efficiency DH5α *E. coli* (Invitrogen™) were used for all plasmid transformations, unless otherwise stated. 5 µL cloning reaction was incubated on ice with 50 µL *E. coli* for 30 minutes. Reactions were heat shocked at 42°C for 20 seconds then returned to ice for 2 minutes. 950 µL S.O.C media (Invitrogen™) was added and samples were incubated at 37°C, 225 rpm for 1 hour. 100µL was spread on LB-agar plates with appropriate antibiotic. Plates were incubated overnight at 37°C.

2.3.2 Plasmid Isolation

Colonies from LB-agar plates were picked with a sterile loop and inoculated into 5 mL or 100 mL LB media with appropriate antibiotics, for miniprep and maxiprep respectively. Cultures were grown overnight at 37°C, 225 rpm.

Cultures were pellet by centrifugation at 3000 rpm for 15 minutes or 5000 xg for 15 minutes, for miniprep and maxiprep respectively. Minipreps were carried out using the peqGOLD Plasmid Miniprep kit (C-line) and maxipreps were carried out using the peqGOLD XChange Plasmid Maxi EF Kit, according to the manufacturer's instructions.

2.4 Tissue Culture Methods

2.4.1 Maintenance of Immortalised cell lines

Handling of all cell lines was carried out in a biological safety cabinet. All cells were grown in a humidified incubator at 37 °C, 5% CO_2. All cells were maintained in 75 cm^2 flasks (Greiner bio-one).

Once cells reached 90% confluence, growth media was aspirated from the flask and cells were incubated with 3 mL TrypLE Express (Life Technologies) until they lost adherence, approximately 5 minutes. 7 mL culture media was added and cells were pipetted several times to remove cell clumps. 1 mL of cell suspension was then transferred to a new flask containing 25 mL fresh media. Cells were then returned to the humidified incubator (37 °C, 5% CO_2).

Cell lines were regularly tested for the presence of mycoplasma infection using a MycoAlert™ mycoplasma detection kit (Lonza). Contaminated cells were disposed of.

2.4.2 Transfection

Cells were seeded according to Table 2.12, 24 hours prior to transfection in 12-well plates (Nunc™). On the day of transfection, 2 µg DNA was mixed with 200 µL OptiMEM (Life Technologies) media and 4 µL peqFECT DNA transfection reagent (peqlab). This mixture was incubated at room temperature for 20 minutes to allow DNA-lipid complexes to form, before being added dropwise to the 12-well plate – 200 µL per well.

Table 2.12 Cell line manipulation. Seeding densities for 12-well plates are indicated.

Cell Line	Seeding Density	Concentration of Puromycin
A549	1.5 x 10^5	5 µg/mL
HEK 293-FT	2.0 x 10^5	2 µg/mL
RPE-1	1.5 x 10^5	5 µg/mL
HCT116	1.5 x 10^5	2 µg/mL
HeLa	1.5 x 10^5	-

For the cell proliferation assays, cells were transfected in 96-well plates (Nunc™). For this, cells were seeded at 1.5 x 10^4 cells per well 24 hours prior to transfection.

Mastermixes for each transfection were used; 200 ng DNA was mixed with 20 µL OptiMEM medium and 0.4 µL peqFECT DNA transfection reagent for each well to be transfected. These were incubated at room temperature, and then 20 µL was added dropwise to each well.

2.4.3 Cell Proliferation Assay

2.4.3.1 Drug treatment

MPA (Sigma-Aldrich®) was reconstituted in DMSO to a stock concentration of 100 mM, and then serially diluted in culture media to the final working concentrations required. Acyclovir monophosphate (Santa Cruz Biotechnology) was reconstituted in tissue culture-grate water to a stock concentration of 100 mM, and then serially diluted in culture media to the appropriate working concentrations. 24 hours post transfection, culture media was aspirated and replaced by 200 µL per well fresh culture media containing appropriate concentrations of drug(s), or vehicle (DMSO) only controls. Cells were then incubated in a humidified incubator at 37 °C, 5% CO_2, for 72 or 120 hours, depending on the experiment.

2.4.3.2 BrdU ELISA

Cell proliferation was determined using a Cell Proliferation ELISA, BrdU (colorimetric) (Roche) according to the manufacturer's instructions. Absorbance was measured at 370 nm (reference wavelength 492 nm) using a plate reader.

2.4.4 Isolation of Single Cell Colonies

2.4.4.1 Puromycin Selection

24 hours post transfection; media was aspirated and replaced with media containing puromycin (Gibco) (Table 2.12). Cells were incubated in the presence of puromycin for 48 hours at 37 °C, 5% CO_2, in a humidified incubator.

2.4.4.2 Fluorescence Activated Cell Sorting

Following 48 hour puromycin treatment, media was aspirated from each well and 0.5 mL TrypLE Express was added. Plates were returned to the incubator for approximately 5 minutes to allow cells to disassociate, and then 0.5 mL media was added to each well and cells were pipetted up and down to remove cell clumps. The

cell suspensions were transferred to 4.5 mL FACS tubes (BD Falcon). Cells were pelleted by centrifugation at 1200 rpm for 5 minutes. Supernatant was aspirated, cells were resuspended in 300 µL PBS and placed on wet ice.

A BD FACSJazz™ machine (BD Biosciences) was used to plate individual cells into each well of a 96-well plate containing 100 µL warmed media. This was carried out by the FACS technician Elisabeth Freyer.

Cells were allowed to grow for up to 7 days, at which point colonies deriving from single cells were identified. These were left to grow to confluence, before being transferred to 48-well plates (Nunc™). Once confluent, these were sequentially transferred to larger and larger culture vessels, up to a 25 cm^2 flask (Greiner bio-one).

2.4.5 Flow Cytometry

Media was aspirated from each well and 0.5 mL TrypLE Express was added. Plates were returned to the incubator for approximately 5 minutes to allow cells to disassociate. Then 0.5 mL media was added to each well and cells were pipetted up and down to remove cell clumps. The cell suspensions were transferred to 4.5 mL FACS tubes (BD Falcon). Cells were pelleted by centrifugation at 1200 rpm for 5 minutes. Supernatant was aspirated, cells were resuspended in 300 µL PBS and placed on wet ice.

A BD LSR Fortessa machine (BD Biosciences) was used to analyse the number of fluorescent cells in each sample. For each experimental sample 50,000 live single cells were analysed.

2.4.6 Genomic DNA Isolation

Cells were removed from wells by incubation with TrypLE Express. Media was added to each well to inhibit the action of the TrypLE Express. Cells were pipetted to make a homogenous solution, and then transferred to 1.5 mL tubes. Cells were pelleted by centrifugation and media aspirated. Cell pellets were stored at -20°C until DNA extraction could be carried out.

Genomic DNA was extracted using the DNeasy Blood and Tissue kit (Qiagen), according to the manufacturer's instructions. DNA was eluted in 100 µL buffer AE.

2.4.7 Total RNA Extraction

Cells were seeded in 6-well plates at a twice the seeding density of a 12-well plate (Table 2.12) and incubated in a humidified incubator for 24 hours at 37 °C, 5% CO_2. Cells were removed from plates and transferred to 1.5 mL tubes as described in Section 2.4.6. Cells were pellet and media removed, pellets were stored at -70°C until RNA extraction could be carried out.

RNA extraction was carried out using RNeasy Plus mini kit (Qiagen®), according to the manufacturer's instructions. RNA was eluted in 30 µL NF-water. RNA samples were then treated with TURBO DNase I to remove any genomic DNA contamination; 2 U of DNase I were added to each sample and incubated at 37°C for 30 minutes. Samples were then purified using spin columns (Qiagen®) and eluted in 30 µL NF-water. RNA concentration and integrity was analysed using an Agilent bioanalyser. Any sample with a RIN < 7 was discarded.

2.4.8 Protein Extraction

Cells were seeded in 6-well plates as described in Section 2.4.7. Cells were harvested and cell pellets made as described above (Section 2.4.7), pellets were stored at -20°C until extraction could be carried out. Pellets were resuspended in 200 µL RIPA buffer supplemented with a protease inhibitor cocktail (cOmplete tablets, Mini EDTA-free EASYpack; Roche) and incubated on a rotating mixer at 4°C for 10 minutes. Samples were centrifuged at top speed at 4°C for 15 minutes, and the supernatant transferred to new 1.5 mL tubes.

2.4.9 Characterisation of Clonal Lines

Target regions were amplified by PCR with the same primers used for the Surveyor assay (Table 2.6). For *EEF1A2*, the following primers were used:

Forward: 5′- TGGTTGAGGAAGGGATCTGG
Reverse: 5′- GATGGGGACGTGGACACA

The primers used to sequence clonal lines generated with SSNs are listed in Table 2.13.

Table 2.13 Sequencing primers for clonal lines

Primer	Sequence
EEF1A1_Ex3_Internal_Seq	5′-AATTAAGGGCTGGGGACAAG
EEF1A1_Ex4_Internal_Seq	5′-CCACCCTACAGCCAGAAGAG
IMPDH2_Internal_Seq _1	5′-GGAGTGTCAAACCAAGTTTCTG
IMPDH2_Internal_Seq _2	5′-TGTTCCTGTGTTGTCCTCCTT
IMPDH2_Internal_Seq _3	5′-AGGAAGGGTGTTTTGGGTGT
EEF1A1_Internal_Seq_F	5′-ACTCTGACACTGGCTGGAT
EEF1A1_Internal_Seq_R	5′-GTTTATCCCATCTGGCGGCT

2.5 Statistical Analyses

All statistical analysis was carried out using Prism 4 (version 4.03; Graphpad). Unpaired T-tests were used to compare between two groups of data. ANOVA with Bonferroni correction was used to compare multiple groups. P-values of >0.05 were considered statistically significant. Data are presented as means ± S.E.M. unless otherwise stated.

2.6 Recipes

2.6.1 General Recipes

20x TBE
216.0 g Tris base
110.0 g Boric Acid
80 mL 0.5 M EDTA, pH 8.5

1x PBS
8 g NaCl
0.2 g KCl
1.44 g Na_2HPO_4
0.24 g KH_2PO_4

PBS-T
1x PBS
0.2% Tween-20

Orange G Buffer
100 mg Orange G
30% Glycerol

1x Surveyor Buffer
10 mM Tris-HCl, pH 8.8
15 mM $MgCl_2$
50 mM KCl

10% Separating Gel (2 Gels)
6.68 mL dH_2O
4 mL 1.5 M Tris, pH 8.8
5.2 mL 30% Acrylamide
80 µL 20% SDS
10 µL TEMED
40 µL AMPS

4.3% Stacking gel (2 Gels)
5.95 ml dH_2O
2.5 ml 0.5M Tris-HCl pH 6.8
1.45 ml 30% Acrylamide
50 µL 20% SDS
5 µL TEMED
50 µL 25% AMPS

Laemmli loading buffer (2x)
60mM Tris-HCl pH 6.8
0.1% Bromophenol blue
10% Glycerol
2% SDS

RIPA Buffer
50 mM Tris pH 7.5
150 mM NaCl
0.5% (w/v) Sodium Deoxycholate
1% (v/v) NP-40 (Igepal)
0.1% (v/v) SDS

Transfer buffer (1x)
700ml dH_2O
200ml Methanol
100ml TGS (10x)

2.6.2 Bacterial Recipes

Luria Broth (LB)
10.0 g Tryptone
5.0 g Yeast Extract
10.0 g Sodium Chloride
1.0 g Glucose

LB-Agar
10.0 g Tryptone
5.0 g Yeast Extract
10.0 g Sodium Chloride
1.0 g Glucose
15.0g Agar

Freezing Media
70% LB
30% Glycerol

2.6.3 Tissue Culture Recipes

HEK 293-FT Growth Media
Dulbecco's Modified Eagle Medium (DMEM)
10% Foetal calf serum (FCS)
0.1 mM Non-essential amino acids
2 mM L-Glutamine
1% Penicillin/Streptomycin

A549 Growth Media
DMEM
10% FCS
1% Penicillin/Streptomycin

HeLa Growth Media
DMEM
10% FCS
1% Penicillin/Streptomycin

RPE-1 Growth Media
DMEM:F12
10% FCS
0.01 mg/mL Hygromycin B
2 mM L-Gluatmine
1% Penicillin/Streptomycin

HCT116 Growth Media
DMEM:F12
10% FCS
2 mM L-Glutamine
1% Penicillin/Streptomycin

Freezing Media
90% FCS
10% DMSO

Chapter 3 Transgene Integration Using TALENs and CRISPR/Cas9

3.1 Introduction

Established methods for transgene integration are inefficient and can result in variable expression due to silencing. Recently, the advent of SSNs has greatly improved the ease and efficiency of specifically targeting transgene integration into predetermined loci, which can be chosen for their ability to permit robust long-term expression. The challenge is to identify such loci and develop effective SSN tools to allow them to be exploited in this way. In this work, I aimed to investigate four human genomic loci, to design and build SSNs to target each locus, and to validate the ability of each site to support transgene integration and robust long-term expression, as a proof of principle for the generation of stable lentiviral producer cell lines. The four sites chosen were the adeno-associated virus (AAV) integration site 1 (AAVS1), the human homologue of the mouse *Rosa26* locus (hROSA26), the inosine monophosphate dehydrogenase 2 (*IMPDH2*) gene, and the eukaryotic translation elongation factor 1 alpha 1 (*EEF1A1*) gene.

The AAVS1 locus located on human chromosome 19 (position 19q13.42), was first identified as a hotspot for AAV integration (Kotin et al. 1992). Hallmarks of open chromatin, specifically a DNase I hypersensitivity site and an insulator, are found here (Ogata et al. 2003) and indicate its potential to support transgene expression. The AAVS1 site occurs within the *PPP1R12C* gene, which encodes protein phosphatase 1 regulatory subunit 12C (PPP1R12C) whose function has not been fully characterized. Other groups have demonstrated the ability of this locus to support transgene integration and expression without observable adverse consequences (Chang & Bouhassira 2012; Wang et al. 2012; Zou et al. 2011; DeKelver et al. 2010; Smith et al. 2008).

The hROSA26 locus located on human chromosome 3 (position 3p25.3) was identified as the human homologue of the mouse *Rosa26* locus (Irion et al. 2007). Both the mouse and human ROSA26 loci express an antisense non-coding RNA; for hROSA26 this is THUMPD3-AS1 whose function is yet to be characterised. It is expressed in a range of adult and embryonic stem cell lines (Irion et al. 2007) and has been shown to support transgene integration and expression even after stem cell differentiation.

IMPDH2 on human chromosome 3 (position 3p21.31) encodes IMPDH2, a key metabolic enzyme involved in guanine nucleotide biosynthesis and is essential for cell proliferation; as such it is ubiquitously expressed. The activity of IMPDH2, and its closely related isoform IMPDH1, can be specifically inhibited via the drug mycophenolic acid (MPA), making this an attractive site for the development of a drug-selectable transgene integration system, discussed further in Chapter 4.

EEF1A1 encodes the eukaryotic translation elongation factor 1 alpha 1 (eEF1A1), and is located on human chromosome 6 (position 6q14). Expression of this gene is almost ubiquitous in normal adult tissue (Cho et al. 2012), and is believed to be expressed in all immortalized cell lines. Its canonical role is delivery of aminoacyl tRNAs (aa-tRNAs) to the A-site of the ribosome during translation, though other moonlighting functions have been identified. These include transformation susceptibility (Tatsuka et al. 1992); protein degradation and protection against apoptosis (Cho et al. 2012); interaction with the cytoskeleton (Condeelis 1995); viral propagation and nuclear transport (Mateyak & Kinzy 2010); neurite outgrowth (Hashimoto & Ishima 2011) and a possible role in ageing (Shepherd et al. 1989). It is not yet known whether these functions are shared by its closely related isoform eEF1A2, which has a reciprocal expression pattern, as appropriate means to probe this have yet to be developed. This is further discussed in Chapter 5.

Currently most clinical-grade lentiviral vectors are produced via transient transfection (Aiuti et al. 2013; Biffi et al. 2013; Cartier et al. 2009; Tebas et al. 2013). Some stable producer lines do exist (Greene et al. 2012; Throm et al. 2009b) but these have generally been created by random integration of viral components followed by selection. SSNs can be utilised to streamline the production process and to make subsequent changes (for example pseudotyping or changing the therapeutic gene) more straightforward.

3.2 Aims

The aim of the work described in this chapter is to develop SSNs – particularly TALENs and CRISPR/Cas9 – to target each of the four chosen loci and use these to investigate the potential of each site to support transgene expression and robust long-

term expression of transgenes in human immortalised cell lines, as a proof of principle for the development of a stable lentiviral producer cell line.

3.3 Results

3.3.1 Investigating Endogenous Transcriptional Activity of Target Loci

To maximize the chances of the target loci supporting robust and long-term expression of transgenes, they should be in transcriptionally active regions – with an open chromatin structure – less susceptible to the silencing effects of heterochromatin. In order to assess the transcriptional activity of each of the target sites, I carried out qPCRs in a panel of immortalised cell lines derived from a range of tissues. To determine the best normalisation strategy, I used the geNorm kit (Primerdesign) to analyse the stability of expression of six reference genes across the panel of immortalised cell lines. From this, I chose five reference genes, which showed the greatest stability: *GAPDH, ATP5B, TOP1, UBC,* and *SDHA*.

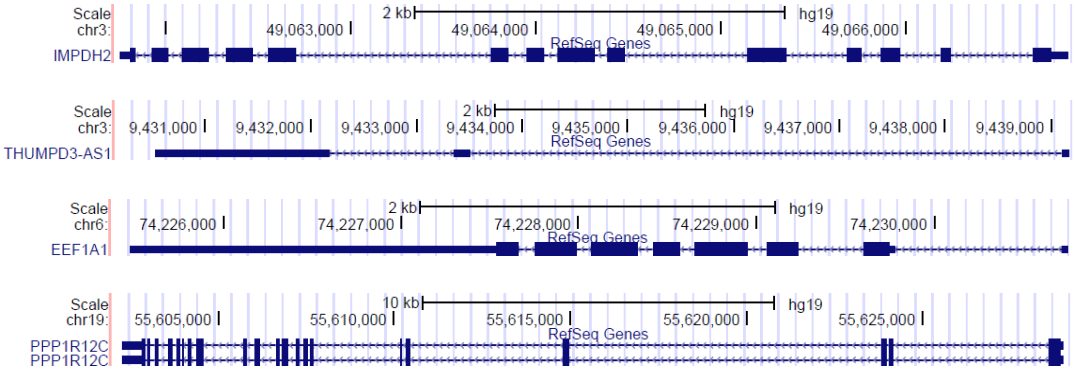
The *IMPDH2* and *EEF1A1* loci encode genes, giving rise to IMPDH2 and eEF1A1 respectively (Figure 3.1), so I was able to monitor transcriptional activity at each locus directly using qPCR. However, as neither the AAVS1 nor hROSA26 loci are genes, I had to use surrogate markers of expression. The AAVS1 locus coincides with the *PPP1R12C* gene (Figure 3.1A); in order to observe transcriptional activity of this locus I designed primers against *PPP1R12C*. The hROSA26 locus contains the open reading frame for a long non-coding RNA THUMPD3-AS1 (Figure 3.1A); I used the expression of this to monitor transcriptional activity at this locus.

Each of the four target loci are transcriptionally active in all cell lines tested (Figure 3.1B). *IMPDH2* shows the highest level of expression (note y-axis), followed by hROSA26, while the least transcriptional activity is seen at the AAVS1 locus, however, *PPP1R12C* shows the most consistent expression across the cell lines tested. These data demonstrate that all loci are transcriptionally active in a range of cells and so have the potential to be useful sites for transgene integration and expression.

This is in line with data from the ENCODE project (Figure 3.1C), which indicates that all target loci lie within transcriptionally permissive regions indicative of open

chromatin. This is illustrated by the presence of acetylated histone H3 at lysine 27 (H3K27Ac). In addition, all loci are associated with multiple DNase I hypersensitivity sites, indicating all loci lie within areas of less compact chromatin.

A



B

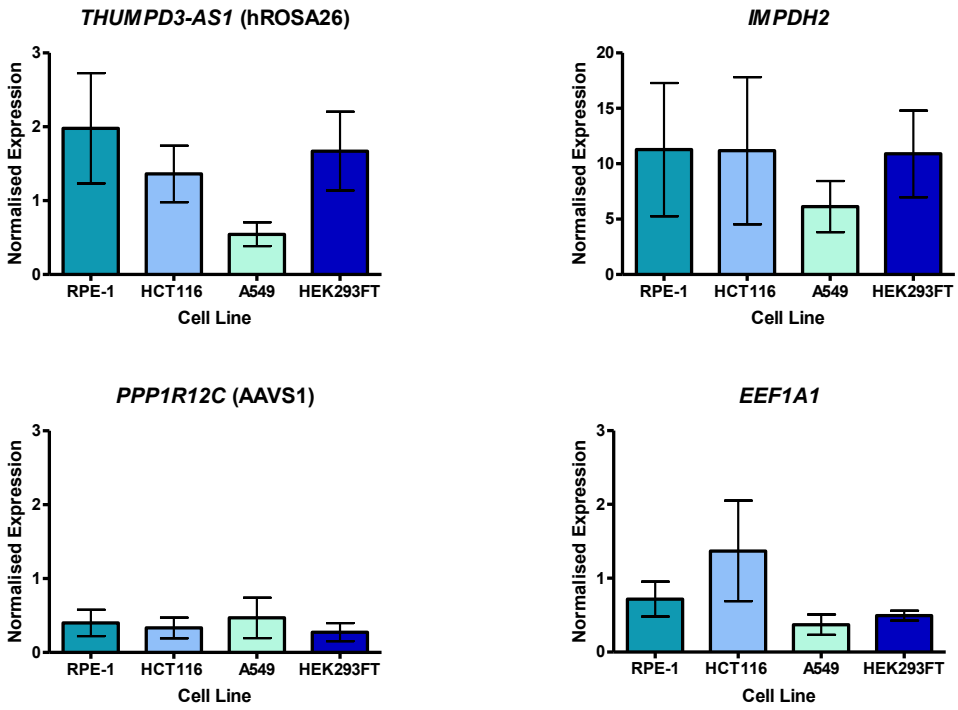


Figure 3.1 Endogenous expression at chosen target sites. Page 1 of 2.

C

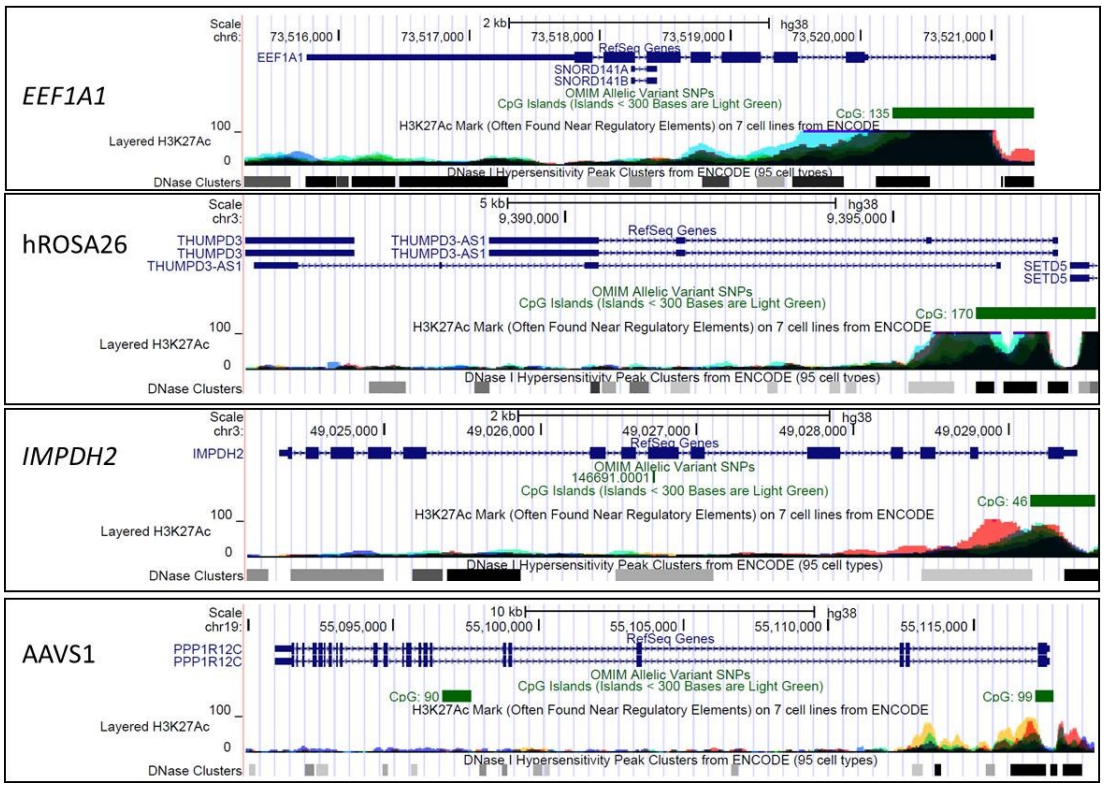


Figure 3.1 Continued, page 2 of 2. A) Genomic organisation of target sites. AAVS1 and hROSA26 are not genes; to monitor expression, the expression of genes in these regions was investigated. Figure generated using USCS genome browser, assembly GRCh37/hg19. B) qPCR expression data for each target locus. Expression of each gene is normalised to the expression of five reference genes (*GAPDH*, *SDHA*, *TOP1*, *UBC*, and *ATP5B*). Data represent the means of three independent experiments. Error bars show S.E.M. C) ENCODE data for each target site, showing location of histone H3K27Ac and DNase I hypersensitivity clusters. Figure generated using USCS genome browser, assembly GRCh37/hg19.

3.3.2 Design and Construction of SSNs

3.3.2.1 TALENs

Due to the relative infancy of this technology, and to act as a positive control for the methodology, I constructed a TALEN pair targeting AAVS1 which had already been shown to be active (Sanjana et al. 2012), alongside a TALEN pair of my own design (targeting the hROSA26 locus).

I used the web-based software devised by Cermak et al. (2011; available at https://tale-nt.cac.cornell.edu/node/add/talen) to identify all potential TALEN target sites within the locus of interest – hROSA26. The input sequence I used was taken from Irion et al. (2007) who first identified the hROSA26 locus. This resulted in 23 potential TALEN target sites. Of these, only eight had no predicated off-target sites; I chose one target site for which I would construct a TALEN pair located downstream of the putative exon one, as defined by Irion et al. (2007), within the first intron of *THUMPD3-AS1*.

I constructed all TALENs using the method devised by the Zhang laboratory (Sanjana et al. 2012) which uses a hierarchical ligation strategy and makes use of the Golden Gate cloning technique (Engler et al., 2009, 2008; Weber et al., 2011). The TALENs I constructed are detailed in Table 3.1.

Table 3.1 Constructed TALENs targeting AAVS1 and hROSA26. Genomic locations are given for entire target sequence including spacer region. Half repeats (encoded by the backbone plasmid) are given in parentheses.

TALEN	Genomic Location	Target Sequence	RVD Sequence
AAVS1_R	Chr19: 55627109-55627158	5′-TGTGGGGTGGAGGGGAC	NG NG HD NG NN NG HD NI HD HD NI NI NG HD HD NG NN (NG)
AAVS1_L		5′-GTCCCCTCCACCCCACA	NN NG HD HD HD HD NG HD HD NI HD HD HD HD NI HD NI (NN)
hROSA26_R	Chr3: 9437057-9437108	5′-GTGTGATAGCTTCCAAAT	NN NG NN NG NN NI NG NI NN HD NG NG HD HD NI NI NI NG (NG)
hROSA26_L		5′-GGATCAAAAGATAACACT	NN NN NI NG HD NI NI NI NI NN NI NG NI NI HD NI HD NG (NI)

3.3.2.2 CRISPR/Cas9

Before I could construct TALENs to target all loci of interest, the CRISPR/Cas9 system was developed (Cong et al. 2013; Jinek et al. 2012; Mali, Yang, et al. 2013). This was shown to be equally active and more easily reengineered. I therefore targeted all remaining loci with this system.

To identify sgRNA target sites I used the software developed by Hsu and colleagues (2013; available at http://crispr.mit.edu/). I constructed all sgRNAs (both individual and nickase pairs) using the protocol devised by the Zhang laboratory (Ran, P D Hsu, et al. 2013). In this method, all components of the system are encoded on a single plasmid (pX459 or pX462), which also encodes a puromycin resistance cassette, allowing for selection of transfected cells.

During the development of their CRISPR/Cas9 system, Mali and colleagues (2013) constructed two sgRNAs that targeted AAVS1 (T1 and T2) at sites which overlapped the target of the TALEN pair I had constructed, as such I continued to use the AAVS1 locus as a control. I acquired the gRNA_AAVS1_T2 plasmid as a gift from George Church (Addgene plasmid #41818). This would also allow me to compare TALEN and CRISPR/Cas9 activity directly, but also to have an externally validated reference with which to compare the activity of sgRNAs of my own design and construction. The Church laboratory CRISPR/Cas9 system is a two-plasmid platform in which a human codon-optimised Cas9 is encoded on one plasmid and the sgRNA is encoded on a second plasmid, which must be co-transfected for the system to be active (outlined by Mali et al., 2013b).

I decided it would be useful to compare the activity of the two systems at additional loci, to control for any local environmental effects. As I already had an hROSA26-targeting TALEN pair, I identified sgRNA target sites in the same region. Fortuitously, there was a predicted nickase pair site, which overlapped the TALEN target site (Table 3.2). These sgRNAs were identified using the nickase pair design tool (available at http://crispr.mit.edu/; Hsu et al., 2013). I made only nickase pairs to ensure greater specificity of targeting.

I constructed six sgRNAs to target the *IMPDH2* locus (Table 3.2), both as individual sgRNAs and as three nickase pairs. These nickase pairs were designed prior to the availability of dedicated software, so were selected as top scoring individual sgRNAs which bound opposite DNA strands within 30-bp of each other. Design considerations for these sgRNAs are described in Chapter 4.

As for the *IMPDH2* locus, I constructed the *EEF1A1*-targeting sgRNAs as both individual guides and as nickase pairs (Table 3.2). Like the *IMPDH2*-targeting sgRNAs, I designed these manually from top scoring single guides. I designed these sgRNAs to have a dual role: in addition to targeting *EEF1A1* as a potential site for transgene integration, they were designed to generate *EEF1A1* knockout cell lines. This is described further in Chapter 5, along with the design considerations for these sgRNAs.

Table 3.2 Constructed sgRNAs. The genomic location and sequence of each constructed sgRNA is listed below for all four genomic loci. In some cases, sgRNAs were constructed for use with either wild type Cas9 or Cas9n (D10A), as indicated by the backbone vector. All backbone vectors also contain a puromycin resistance gene to allow for selection of successfully transfected cells.

Name	Target Sequence	Genomic Location	Backbone
sgRNA_hROSA26_92A	5′-GATCCATCTTACATCTTGTTAGG	Chr3: 9437039-9437061	pSpCas9n-2A-Puro
sgRNA_hROSA26_92B	5′-CAATTTGGAAGCTATCACACAGG	Chr3: 9437089-9437111	pSpCas9n-2A-Puro
sgRNA_IMPDH2_1+	5′-AAGCACTGCCCCTACCTTAGGGG	Chr3: 49064904-49064926	pSpCas9-2A-Puro / pSpCas9n-2A-Puro
sgRNA_IMPDH2_1-	5′-CTTTTAAACCCCTAAGGTAGGGG	Chr3: 49064912-49064934	pSpCas9-2A-Puro / pSpCas9n-2A-Puro
sgRNA_IMPDH2_2+	5′-CCCTCTTACACTATGGAATAAGG	Chr3: 49064838-49064860	pSpCas9-2A-Puro / pSpCas9n-2A-Puro
sgRNA_IMPDH2_2-	5′-TCCTTATTCCATAGTGTAAGAGG	Chr3: 49064839-49064861	pSpCas9-2A-Puro / pSpCas9n-2A-Puro
sgRNA_IMPDH2_3+	5′-ATTACAGGGGTGAAGTTCTATGG	Chr3: 49064698-49064720	pSpCas9-2A-Puro / pSpCas9n-2A-Puro
sgRNA_IMPDH2_3-	5′-TGGGTGTGAATGTATGAACGTGG	Chr3: 49064725-49064747	pSpCas9-2A-Puro / pSpCas9n-2A-Puro
sgRNA_EEF1A1_Ex3+ve	5′-GGAGCCCTTTCCCATCTGTAAGG	Chr6: 74229225-74229247	pSpCas9-2A-Puro / pSpCas9n-2A-Puro
sgRNA_EEF1A1_Ex3-ve	5′-CGACTCTTAATCCTTACAGATGG	Chr6: 74229236-74229258	pSpCas9-2A-Puro / pSpCas9n-2A-Puro
sgRNA_EEF1A1_Ex4+ve	5′-AGACCATTGTTAAAAAGCTCTGG	Chr6: 74228618-74228640	pSpCas9-2A-Puro / pSpCas9n-2A-Puro
sgRNA_EEF1A1_Ex4-ve	5′-AAGCCACTTACGTTAGCACTTGG	Chr6: 74228644-74228666	pSpCas9-2A-Puro / pSpCas9n-2A-Puro

3.3.3 Validation of TALEN Activity *in vitro*

To validate the ability of the TALEN pairs to bind and cleave their targets, I carried out an *in vitro* cleavage assay. I generated TALEN proteins *in vitro* using the T_NT® quick-coupled transcription/translation kit (Promega). This yielded good expression of the TALEN proteins (Figure 3.2A & B); both hROSA26 and AAVS1 TALENs were produced at a similar level to the luciferase controls. There was an additional band in the hROSA26 left TALEN lane (Figure 3.2A) which is not present in any other lane. I carried out a luciferase assay (data not shown), which confirmed the activity of proteins produced *in vitro*.

I next carried out the *in vitro* cleavage assay using a PCR-generated target, which I incubated with the TALEN proteins. Figure 3.2C illustrates the activity of the hROSA26 TALEN pair: following agarose gel electrophoresis of the *in vitro* cleavage reaction, cleavage products of predicted size are present (Figure 3.2C, red arrows) and only when the PCR product was incubated with both left and right TALEN proteins. An additional band appears to be present when the right TALEN is present alone though the smaller cleavage band is absent. However, I was unable to show activity for the AAVS1-targeting TALEN pair (Figure 3.2D), which had previously been shown to be active (Sanjana et al. 2012).

Due to the nature of this assay it cannot indicate how the TALENs will act under physiological conditions: because of this and the fact I was unable to detect activity of the AAVS1-targeting TALEN pair, I sought to validate TALEN activity *in cellulo*.

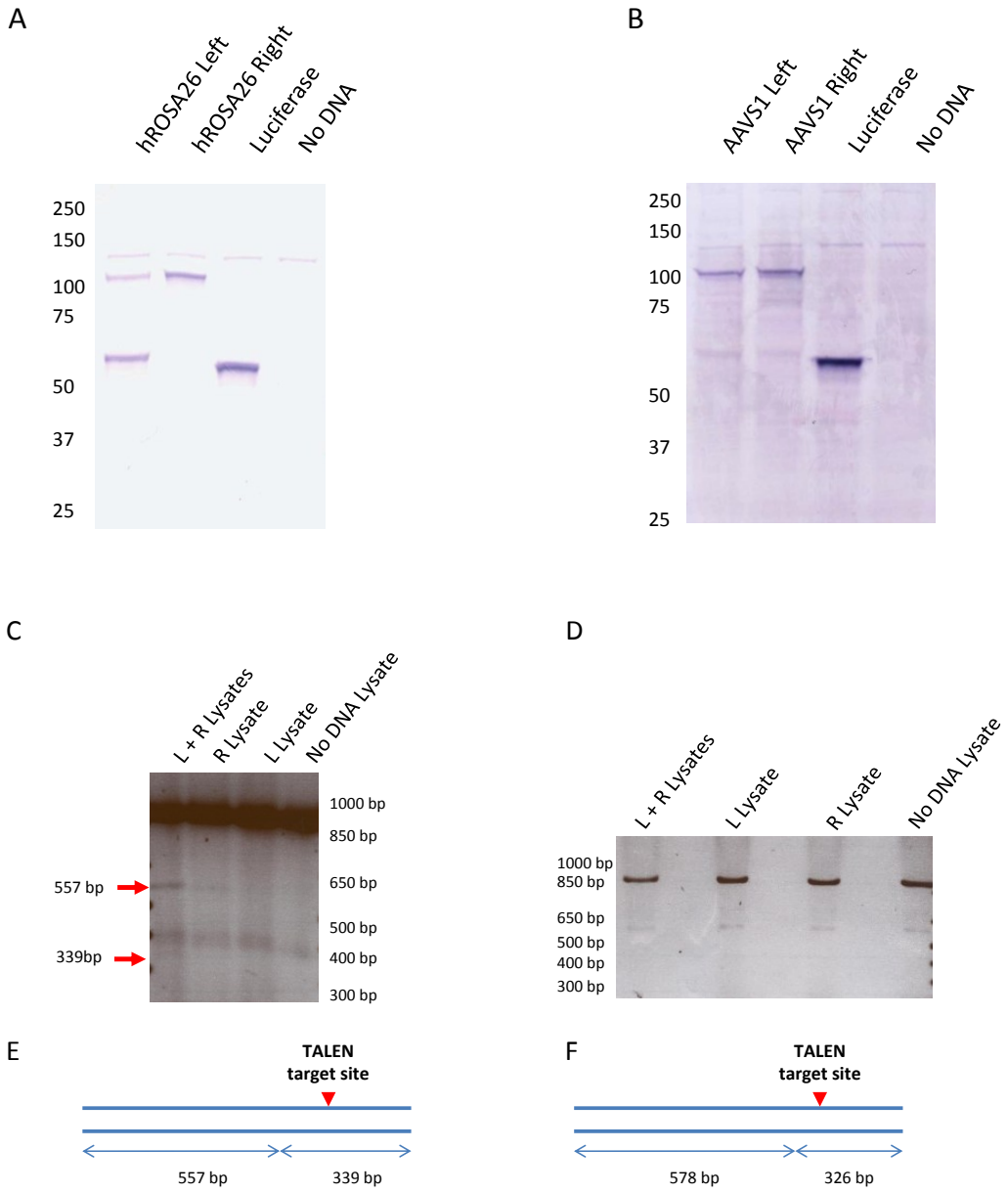


Figure 3.2 Investigating TALEN activity *in vitro*. A & B) TALEN proteins were produced *in vitro* using the TNT Quick Coupled Transcription/Translation Kit (Promega). To allow proteins to be visualised, Biotin-Lysyl-tRNAs were used, so that the resulting proteins contained biotinylated lysine residues. Proteins were analysed by SDS-PAGE, transferred to membranes and visualised using Western Blue® Stabilized Substrate for Alkaline Phosphatase. A) hROSA26-targeting TALENs. B) AAVS1-targeting TALENs. C) *in vitro* cleavage by hROSA26 TALENs, cleavage bands of predicted size are present (red arrows) only when both left and right TALEN proteins are present. D) *in vitro* cleavage by AAVS1 TALENs, the predicted cleavage products are not present. E) Schematic of predicted cleavage sites in hROSA26 target. F) Schematic of predicted cleavage sites in AAVS1 target.

3.3.4 SURVEYOR Assay: Validation of SSN Activity *in cellulo*

I investigated the activity of each SSN following transfection of HEK293FT cells, as these cells are customarily used in the production of lentiviral vectors and express the SV40 large T antigen, which is required for lentiviral production (Gama-Norton et al. 2011). I carried out the SURVEYOR assay 24 hours post transfection for all SSNs.

Figure 3.3 confirms activity of both the TALEN pair I constructed and the sgRNA targeting AAVS1, demonstrated previously (Mali, Yang, et al. 2013; Sanjana et al. 2012). It is interesting to note that I observed 19% indel formation with the AAVS1-targeting TALEN pair (Figure 3.3A) compared with 2.5% observed by Sanjana and colleagues. In contrast, the indel percentage I observed for the sgRNA was similar to that observed by Mali et al. (2013) who used next generation sequencing (NGS) to more accurately determine activity.

I had already demonstrated the potential for the hROSA26-targeting TALEN pair to directly bind and cleave their target (Figure 3.2C). However, I wanted to demonstrate this under physiological conditions.

Figure 3.4A demonstrates that the hROSA26-targeting TALEN pair is active *in cellulo*, confirming the *in vitro* results. In addition, there is no evidence here of the right TALEN acting as a homodimer, under physiological conditions.

I next compared the activity of the TALEN pair and the CRISPR nickase pair, which both target the same site within the hROSA26 locus. Figure 3.4B demonstrates the activity of both TALENs and CRISPR/Cas9 systems at this site. The CRISPR nickase pair resulted in three cleavage products, following agarose gel electrophoresis of the SURVEYOR reaction, instead of the expected two products.

Figure 3.5A shows almost no detectable activity was observed for any of the individual sgRNAs targeting *IMPDH2*; minor activity is seen with sgRNA_1-ve (red arrows).

In addition, there is little evidence of activity for most of the *EEF1A1*-targeting sgRNAs after 24 hours (Figure 3.6). However, sgRNA_EEF1A1_Ex3+ve in combination with wild type Cas9 shows robust activity (Figure 3.6A).

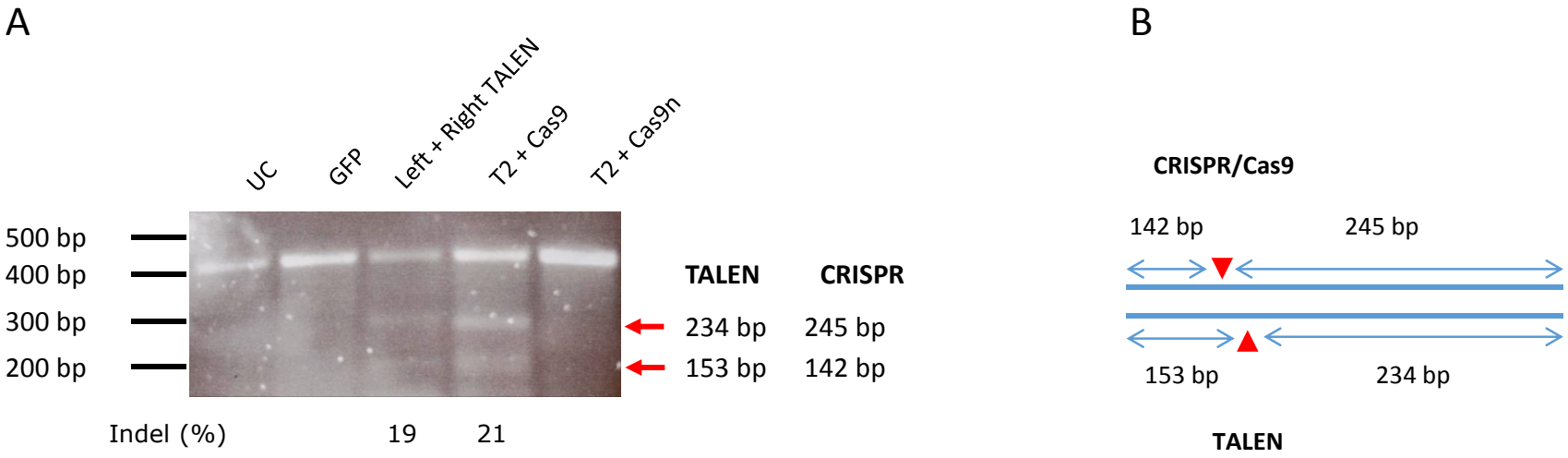


Figure 3.3 Activity of AAVS1-targeting SSNs. A) SURVEYOR assay for AAVS1-targeting SSNs. Red arrows indicate cleavage products. Sizes of predicted cleavage products are indicated. B) Schematic of predicated cleavage products for both TALEN and CRISPR/Cas9 systems. UC = untransfected control.

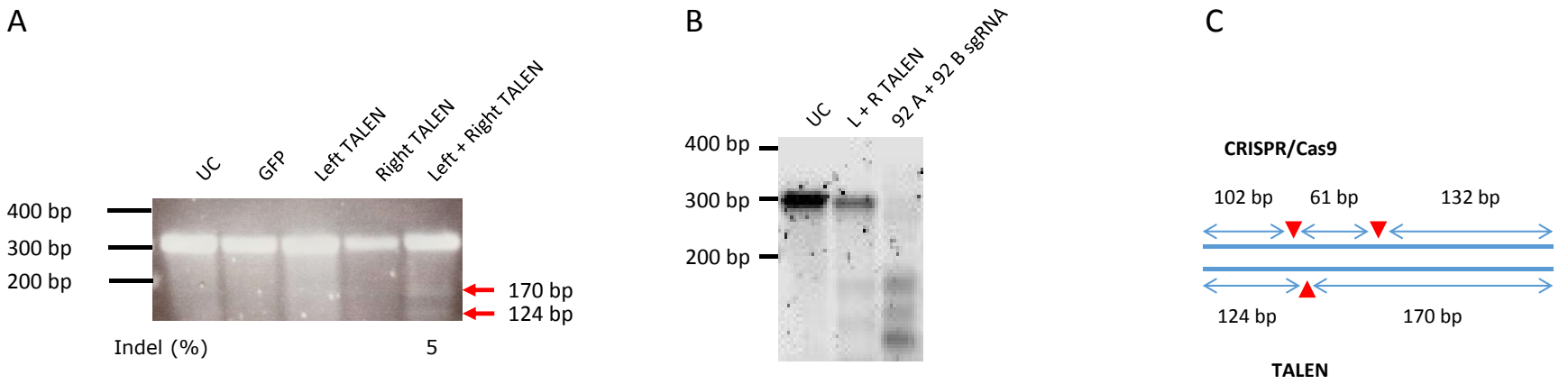


Figure 3.4 hROSA26-targeting SSNs SURVEYOR assay. A) SURVEYOR assay results for hROSA26-targeting TALENs. Red arrows indicate cleavage products, predicted sizes are indicated. B) Comparison of the activity of CRISPR/Cas9 and TALENs targeting hROSA26. C) Schematic of predicted cleavage products for both CRISPR/Cas9 and TALENs at the hROSA26 locus. UC = untransfected control.

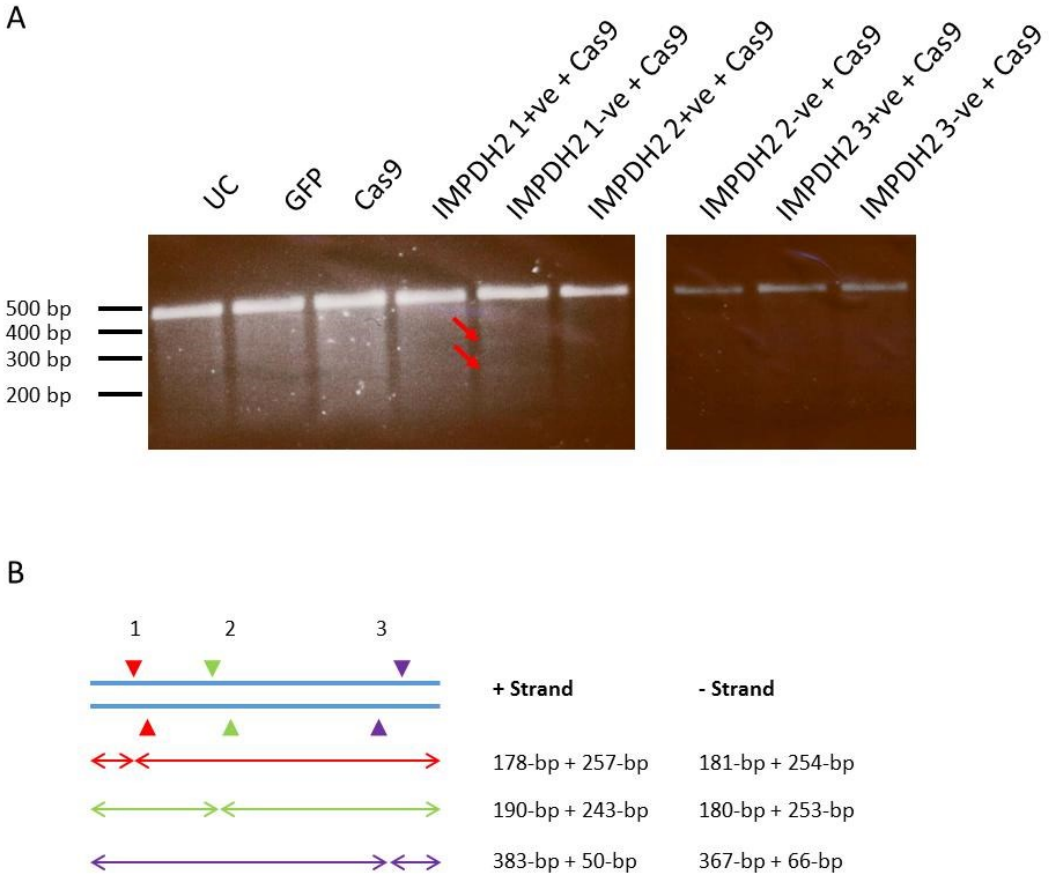


Figure 3.5 SURVEYOR assay for *IMPDH2*-targeting sgRNAs. A) SURVEYOR assay results. Red arrows indicate potential cleavage bands. B) Schematic of predicted cleavage products for individual sgRNAs at the *IMPDH2* locus.

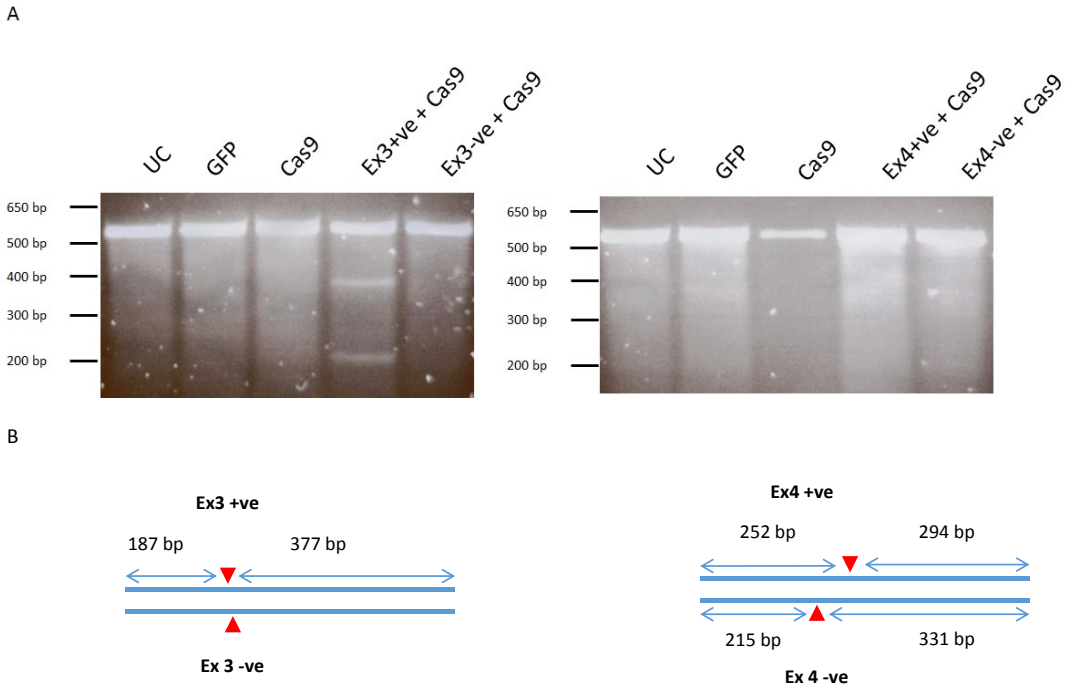


Figure 3.6 SURVEYOR assay for *EEF1A1*-targeting sgRNAs. A) SURVEYOR assay results. Only sgRNA_EEF1A1_Ex3+ve demonstrates any observable activity. B) Schematic of predicted cleavage sites at both sgRNA target sites within the *EEF1A1* locus. UC = untransfected control.

3.3.4.1 Effect of Puromycin Selection on SURVEYOR Assay Outcome

I sought to improve the effectiveness of retrieving cells containing targeted indels at *IMPDH2* and *EEF1A1* by selecting for transfected cells with puromycin. I treated cells for 48 hours with puromycin following transfection; plasmids encoding the CRISPR/Cas9 components also encode a puromycin resistance cassette. After selection, I carried out the SURVEYOR assay as before.

Figure 3.7 demonstrates that all individual sgRNAs (in combination with wild type Cas9) targeting both *EEF1A1* and *IMPDH2* are active following puromycin treatment. However, the nickase pairs targeting *IMPDH2* still do not show convincing activity; pair 1 does show potential activity, as demonstrated by the presence of faint bands (Figure 3.7A, red arrows). On the other hand, both of the nickase pairs targeting *EEF1A1* show robust activity, with no apparent difference between head-to-head and tail-to tail oriented pairs.

In order to better understand the effect of puromycin, I repeated the experiment for the *IMPDH2*-targeting sgRNAs, one batch I treated with puromycin for 48 hours (24 hour post-transfection), and the other I did not treat with puromycin. In both cases, I carried out the SURVEYOR assay 72 hours post transfection

Figure 3.8 demonstrates that activity is observed for all individual *IMPDH2*-targeting sgRNAs at 72 hours post transfection irrespective of puromycin treatment. However, puromycin selection improved the observed indel percentage in most cases; therefore, I carried out all subsequent experiments using 48 hour puromycin selection, 24 hours post transfection.

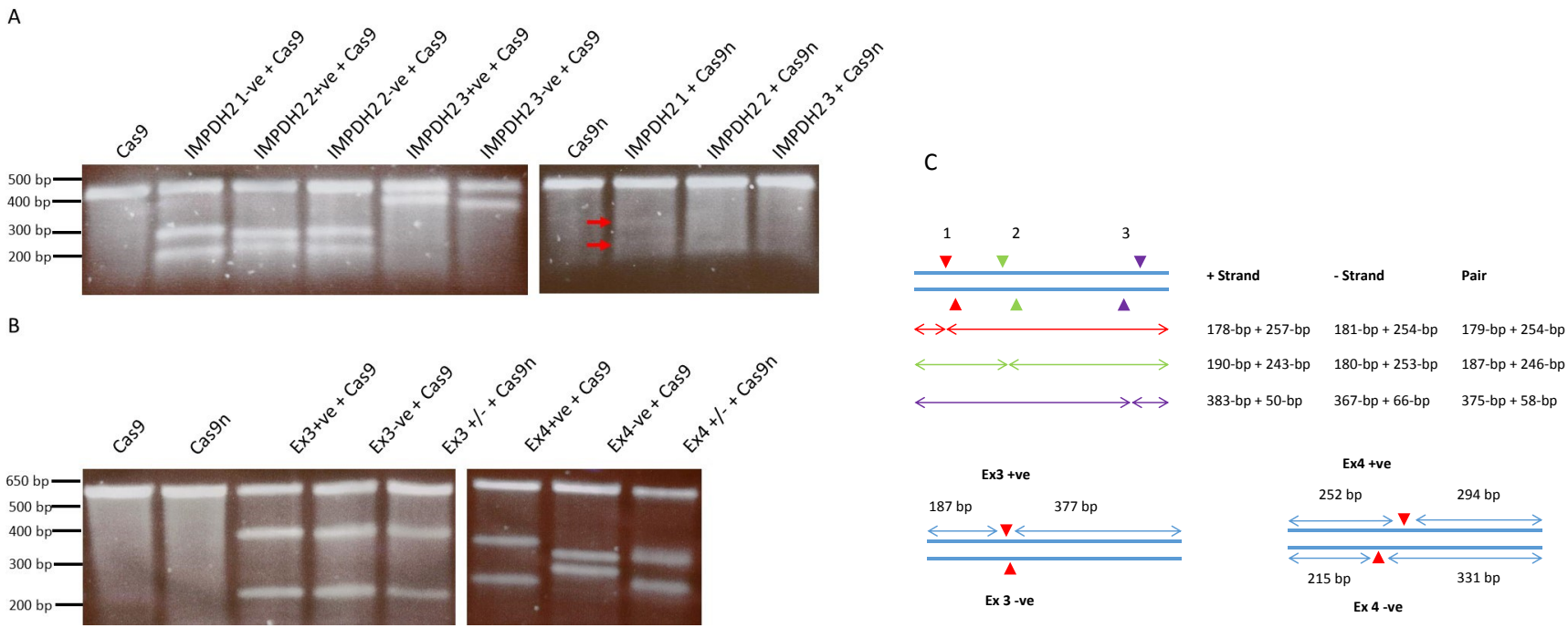


Figure 3.7 SURVEYOR assay post puromycin selection. A) *IMPDH2*-targeting sgRNAs. All single sgRNAs show activity post puromycin selection. There is limited evidence of activity for any of the paired sgRNAs. Red arrows indicate potential cleavage products. B) *EEF1A1*-targeting sgRNAs. All single and paired sgRNAs demonstrate activity following puromycin selection. C) Schematic of predicted cleavage products for all single and paired sgRNAs.

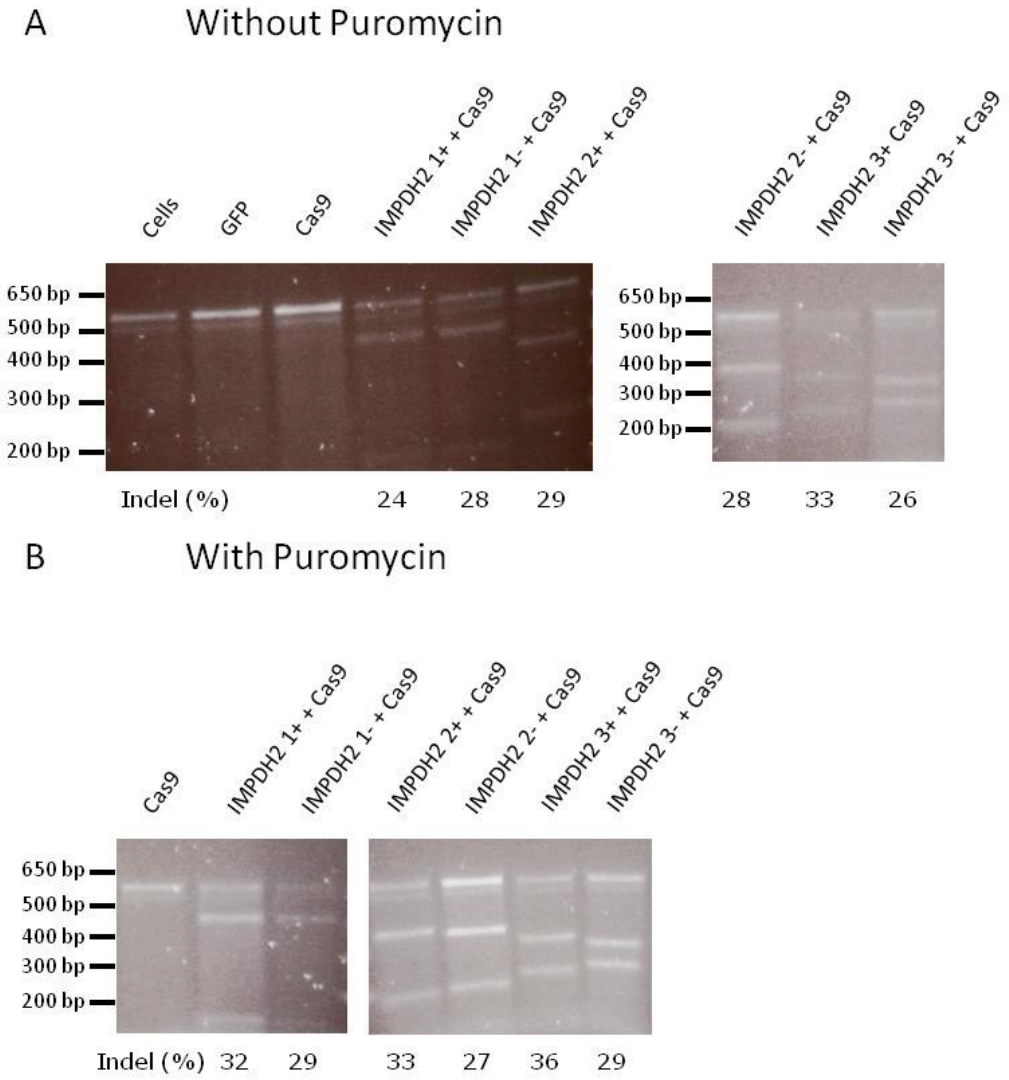


Figure 3.8 SURVEYOR assay ± puromycin selection. Cells were transfected with *IMPDH2*-targeting sgRNAs, and incubated ± puromycin, for a total of 72 hours post transfection. A) SURVEYOR assay on samples not treated with puromycin. B) SURVEYOR assay on samples treated with puromycin. C) Schematic of predicted cleavage products for all single sgRNAs.

3.3.4.2 Assessing SSN Activity in Other Immortalised Cell Lines

After validating the activity of these SSNs in HEK293FT cells, I wanted to investigate if this was consistent in other immortalised cell lines, and investigate whether cell line specific differences could be observed. However, if SNPs are present at the target site, or within the amplified region, it is almost impossible to distinguish these from SSN activity using the SURVEYOR assay, as exemplified by the *EEF1A1* locus in Figure 3.9.

Figure 3.9A shows that for sgRNA_EEF1A1_Ex3+ve and sgRNA_EEF1A1_Ex3-ve, activity cannot be determined by the SURVEYOR assay; the untransfected control lane also has cleavage bands present, which are indistinguishable from the cleavage bands indicative of SSN activity (Figure 3.9B). This is due to the presence of two SNPs in the amplified region: rs4708060 and rs2073466 (Figure 3.10). I confirmed by sequencing that A549 cells are heterozygous for presence of rs4708060 and RPE-1 cells are heterozygous for rs2073466 (Figure 3.10). I also confirmed that SNPs were present at a number of the other target loci; therefore it was not practical to continue to use the SURVEYOR assay, as the presence of SNPs interfere with the interpretation of the results at this particular locus.

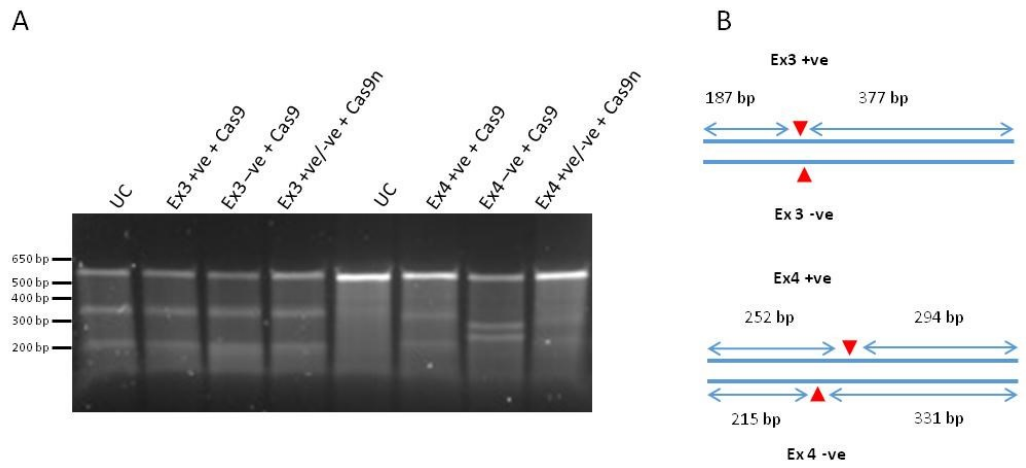


Figure 3.9 Activity of *EEF1A1* targeting sgRNAs in A549 cells. A) SURVEYOR assay results. All samples, even the untransfected control, have the predicted cleavage bands present for Ex3 sgRNAs. B) Schematic of predicted cleavage products.

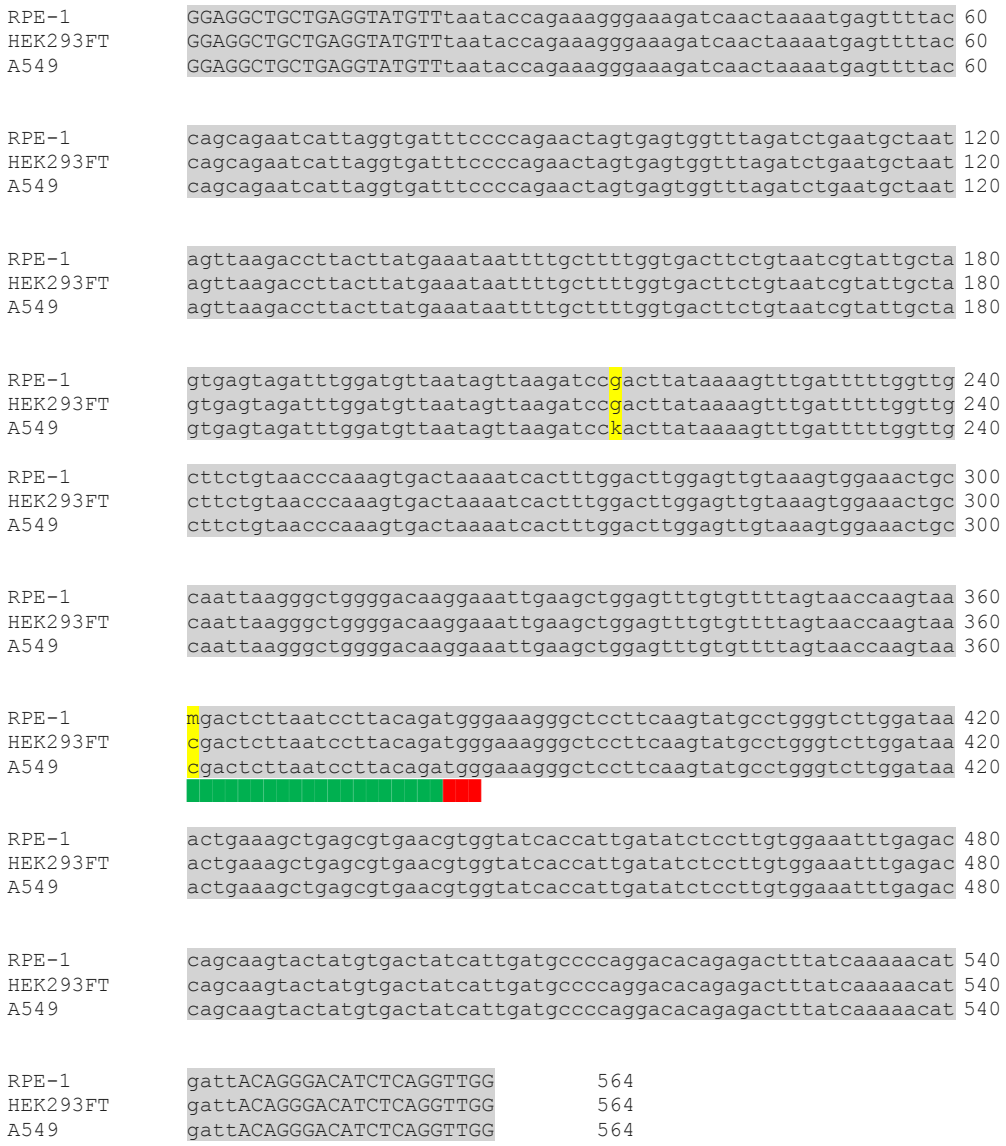


Figure 3.10 Multiple sequence alignment of PCR amplified genomic region from HEK293FT, A549 and RPE-1 cell lines. SNPs are highlighted in yellow. Binding site for sgRNA_EEF1A1_Ex3-ve is indicated by the green box with the PAM highlighted red. Primer sequences for amplified region are given in capitals.

3.3.5 SSN-Driven Transgene Integration

I had shown that it was possible to generate indels, following NHEJ repair of the DSBs, at all my target loci. I next wanted to assess whether I could use these SSNs to target transgene integration and if these loci could support robust and long-term expression of the integrated transgenes.

3.3.5.1 Integration at AAVS1

Both the TALEN pair and the sgRNA targeting AAVS1 can drive HR repair using a short repair template (90mer oligonucleotide) at a modified endogenous locus to repair an *EGFP* gene disrupted by a stop codon and a 68-bp genomic fragment from the AAVS1 locus (Mali et al., 2013). In addition, the sgRNA has been shown to drive transgene integration at the native AAVS1 locus (Mali et al., 2013). I wanted to confirm these findings and test the ability of the TALEN pair to drive integration of a large DNA fragment at the native AAVS1 locus. I therefore constructed a repair cassette containing an *EGFP* gene under the control of the CMV promoter flanked by arms of homology (approximately 800-bp each) to AAVS1 (Figure 3.12A). This cassette was co-transfected into cells with either the TALENs or CRISPR/Cas9 system (T2 sgRNA and Cas9), to test the ability of each system to drive transgene integration at this locus. In addition, I included the sgRNA with Cas9n to test if a single strand nick could drive integration of a large transgene. I had intended to monitor integration via EGFP expression using flow cytometry (Figure 3.11). This proved to be imprecise with a heterogeneous cell population and so I carried out a PCR-based method for transgene integration detection. This relies on locus specific and transgene specific primers, so that a product of 1778-bp would be produced only if correct integration occurred.

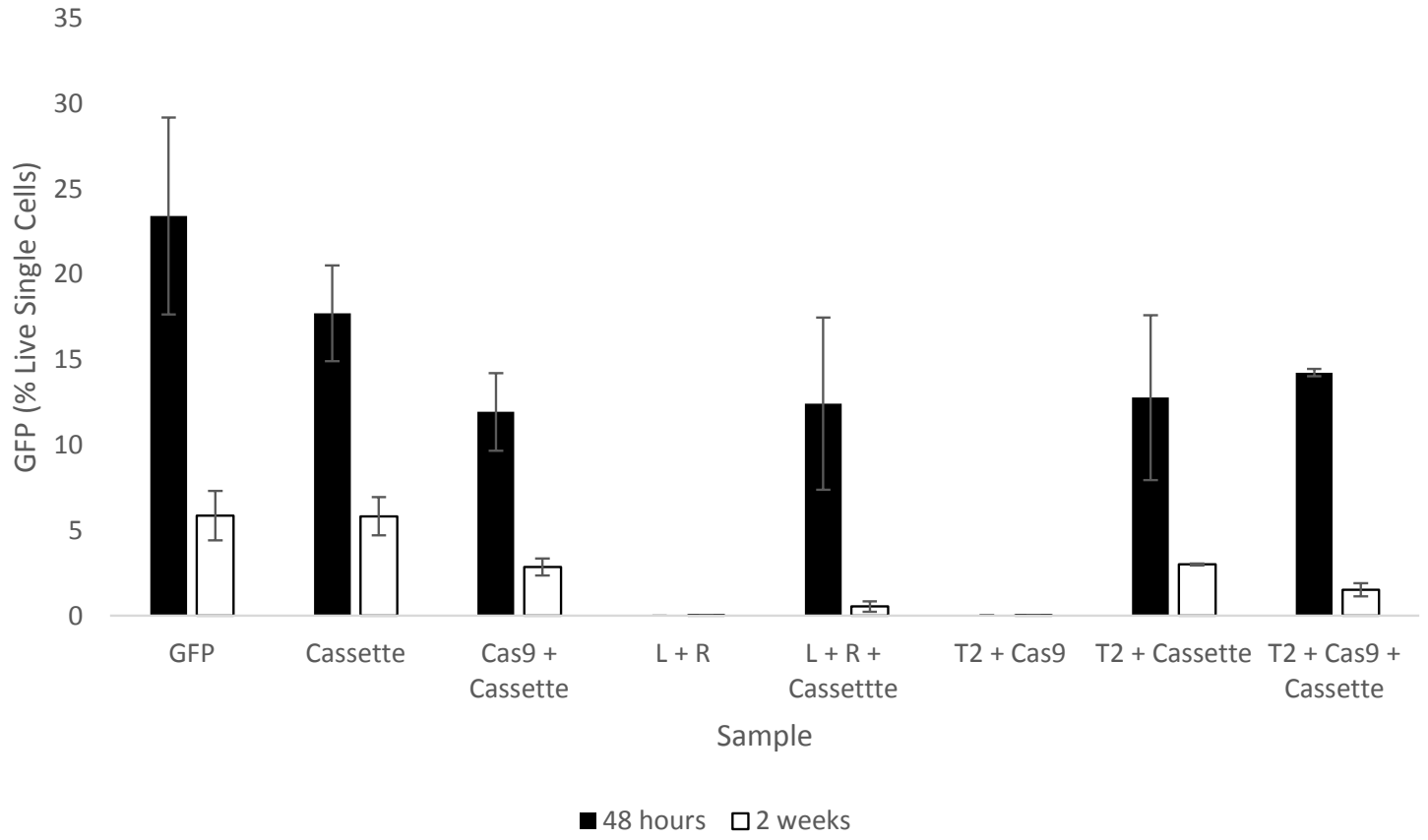


Figure 3.11 EGFP expression in heterogeneous cell populations. EGFP expression was monitored at 48 hours and 2 weeks post transfection. ALL samples showed a rapid decline in EGFP expression. There appears to be no apparent trend in the level of EGFP expression and the presence of SSNs. Data represent the means of two independent samples, error bars indicate standard deviation.

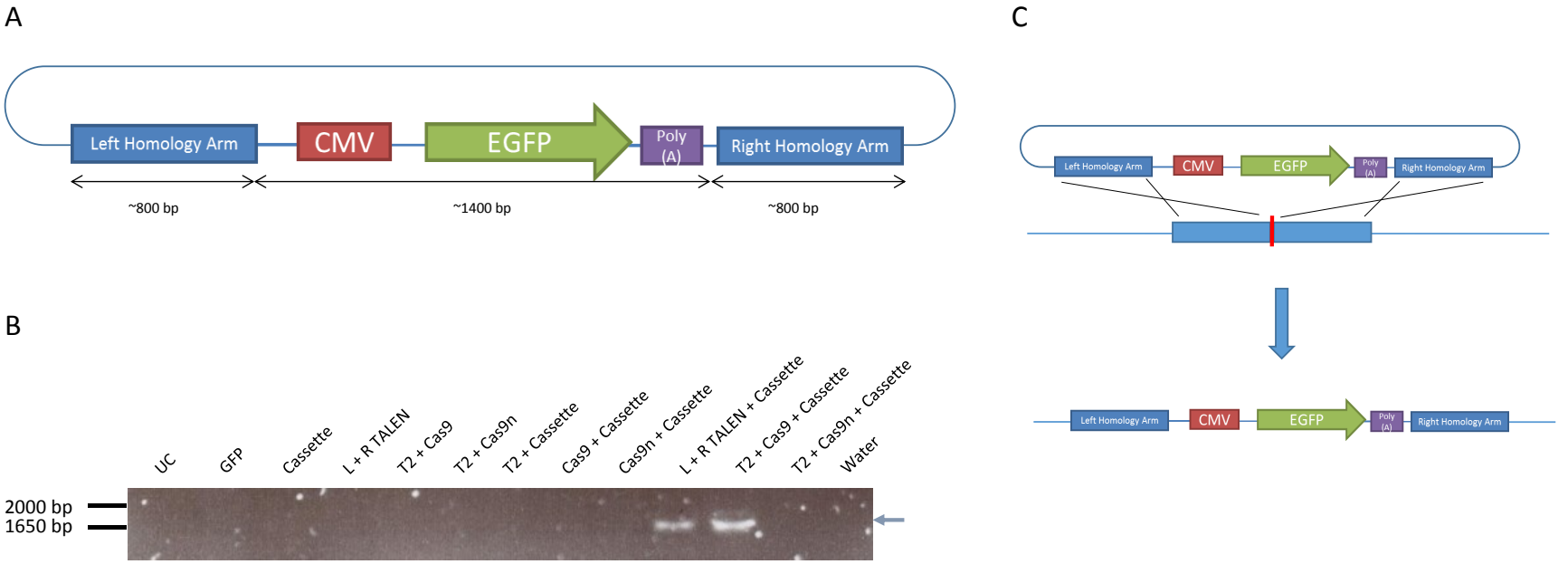


Figure 3.12 Transgene integration at AAVS1. A) Diagram of AAVS1 repair cassette. The cassette contains ~1.6-Kb total homology. The *EGFP* gene is driven by the CMV promoter and contains an SV40 poly A signal. B) Agarose gel confirming the presence of integrated cassette. Primers specific to the AAVS1 locus and the cassette were used to detect the presence of correctly inserted transgenes. Blue arrow indicates PCR product of predicted size. C) Schematic representation of integration of the cassette at the AAVS1 locus. Following cleavage by the SSN. The DSB is repaired by HR using the cassette as a donor template, thus leading to the insertion of the *EGFP* gene. UC = untransfected control.

Figure 3.12B shows that both TALENs and CRISPR/Cas9 can drive transgene integration, but Cas9n cannot. This confirms the results of Mali and colleagues (2013) and demonstrates that TALENs can also drive integration of larger DNA fragments into endogenous loci. It also shows that integration is dependent on DNA DSB formation, as no targeted integration (at this locus) is observed when the cassette alone is transfected into cells. Therefore, both systems can drive the site-specific integration of large (~1.6-Kb) transgenes.

3.3.5.2 Integration at hROSA26

Both the TALEN pair and sgRNA which I used to target AAVS1 were designed by others (Mali, Yang, et al. 2013; Sanjana et al. 2012) and the sgRNA had previously been shown to facilitate the integration of transgenes (Mali, Yang, et al. 2013). In addition, several commercially available transgene integration systems based around AAVS1 had been developed (Sigma-Aldrich® Cat. No. CTI1; GeneCopoeia Cat. No. SH-AVS-K100). I therefore focused on the hROSA26 region, which had been less intensively studied. I also had both a TALEN pair and a double nickase pair of sgRNAs targeting an overlapping site at this locus, allowing for direct comparison of the two systems.

I constructed a transgene cassette by cloning PCR-amplified homology arms (~1.8Kb total homology) either side of an *mCherry* reporter gene under the control of the EF1α promoter (Figure 3.13), which has previously been shown to drive higher transgene expression than the CMV promoter at the *Rosa26* locus: this is particularly apparent when positioned in a sense orientation (Chen et al. 2011), which this cassette was designed to achieve. In addition, the EF1α promoter has been shown to be less prone to silencing than CMV (Teschendorf et al. 2002).

I analysed HEK293FT cells by flow cytometry following transient transfection with and without puromycin treatment, to validate mCherry expression and test if this could be enriched by puromycin selection.

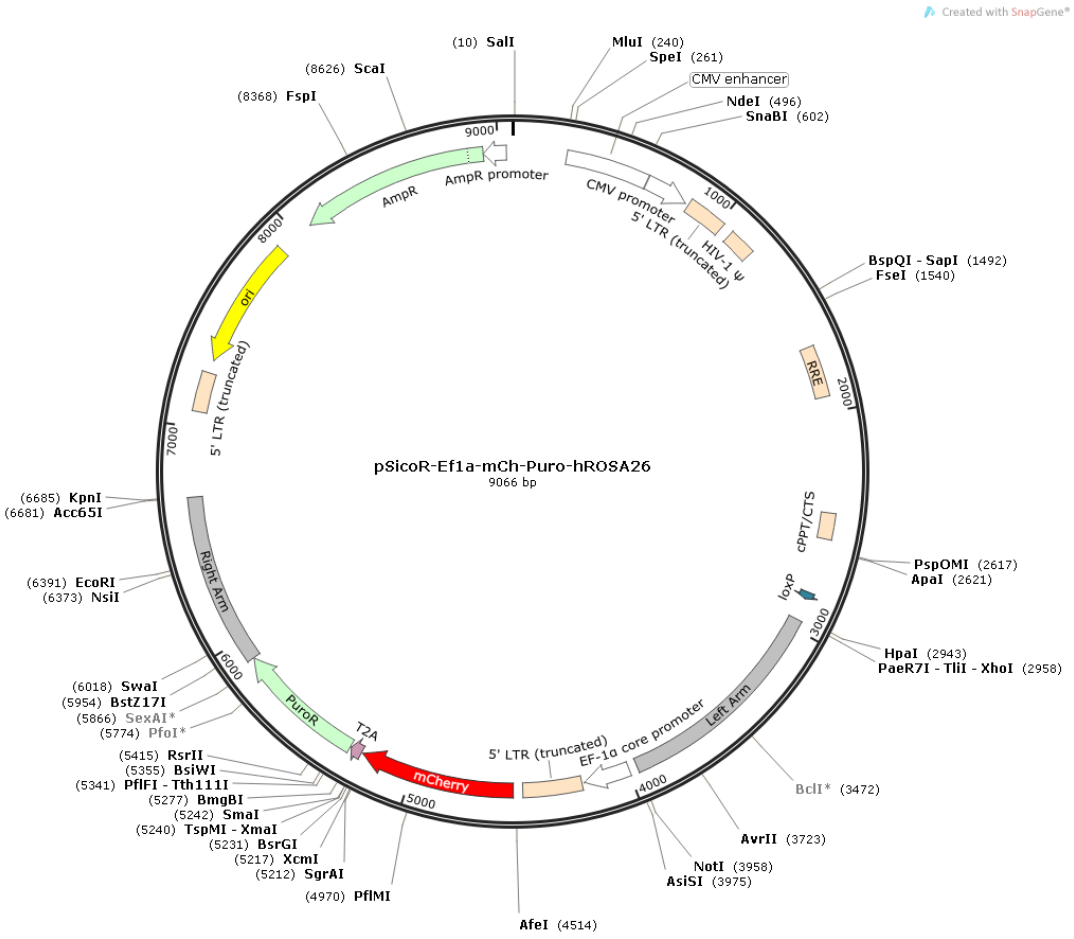


Figure 3.13 Diagram of hROSA26-targeting cassette. The cassette contains ~1.8-Kb total homology. This cassette makes use of the *mCherry* reporter gene driven by the EF1α promoter. This is separated from a puromycin resistance gene (PuroR) by a T2A "self-cleaving" peptide sequence. This results in the production of the two proteins from a single mRNA.

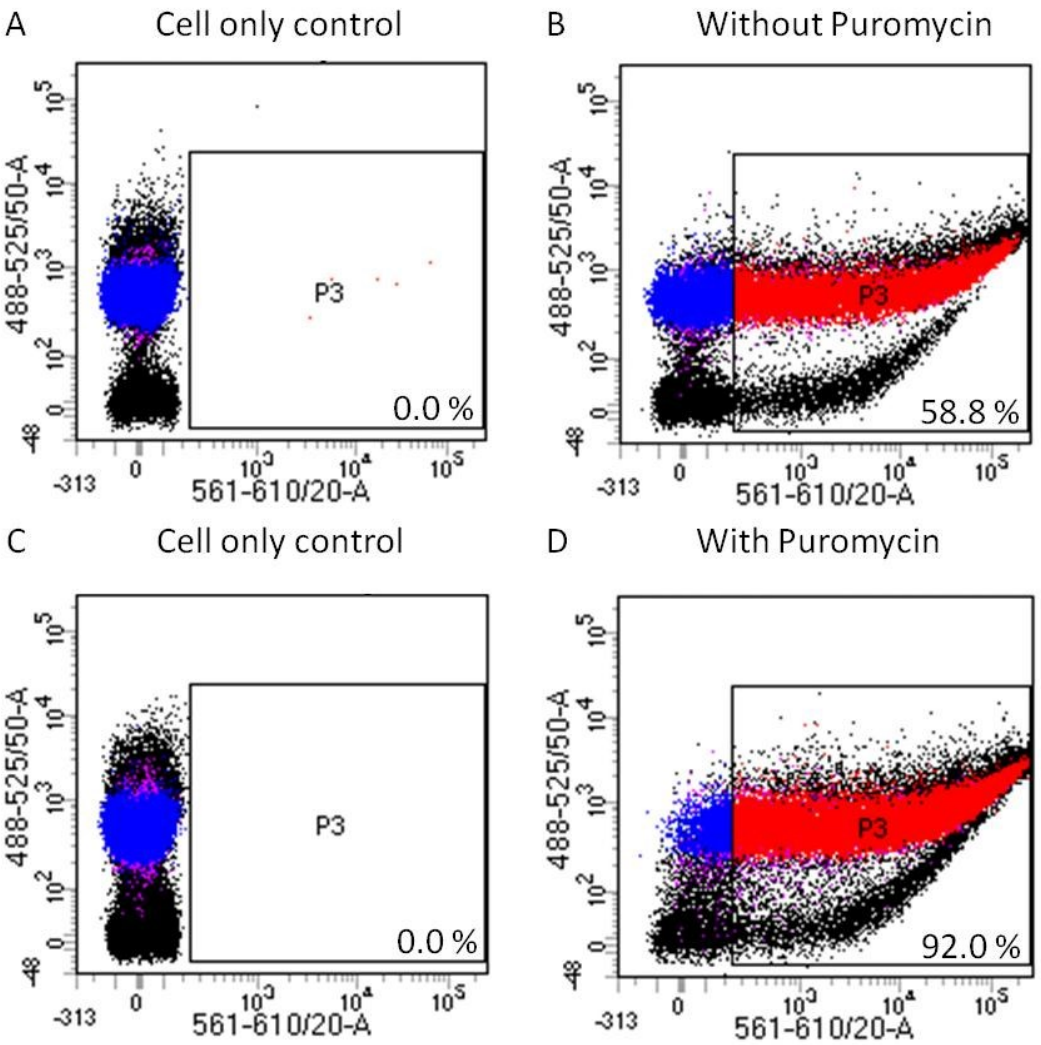


Figure 3.14 Flow cytometric data following transient transfection of cassette, with and without puromycin selection. Scatter plots show mCherry expression levels indicated as percentage of live single cells. Red dots indicate mCherry positive, live single cells; blue dots are live cells; purple dots are live single cells; black dots are dead cells, or cell clumps. For each sample, 50,000 live single cells were analysed. A) Cell only control. B) mCherry with puromycin selection. C) Cell only control. D) mCherry without puromycin selection.

Figure 3.14 demonstrates that a high proportion of cells (58.8% of live single cells) expressed mCherry following transient transfection. In addition, transient puromycin treatment was able to enrich the number of mCherry positive cells (92% of live single cells).

I next sought to ascertain if the hROSA26-targeting TALENs and CRISPR/Cas9 nickase pair were able to drive integration of this cassette. Following transfection of the TALEN or CRISPR/Cas9 machinery with the cassette, I monitored mCherry expression over time using flow cytometry (Figure 3.15).

Initially, the proportion of mCherry positive cells in each population varied quite considerably and expression was rapidly lost over time in all samples (Figure 3.15). At 12-days post transfection, I sorted for mCherry expression in each population by FACS. I then allowed these populations to propagate to the 6-well plate stage before sorting for individual mCherry positive cells to isolate clonal lines again by FACS based on mCherry expression. I was able to generate 18 clonal lines: six lines from an mCherry control, one line from cassette alone, one line from TALENs and cassette and ten lines from CRISPR/Cas9 and cassette. Following a further month of culturing, I analysed mCherry expression in each clonal line and employed a PCR-based method to identify correct integration events.

I detected high levels of mCherry expression in 2/18 (11.1%) clonal lines, one of which was generated with CRISPR/Cas9 and cassette: the other was an mCherry control line. In addition, 6/18 (33.3%) lines showed low-level mCherry expression (mean 17.9% ± 19.9). However, I was unable to detect correct integration by PCR in any of the 18 clonal lines generated (Figure 3.16B), so that in this experiment on-target integration efficiency was 0% (0/11) of lines generated with SSNs.

In investigating the causes of this disappointing outcome, I discovered that the cassette's homology arms had inadvertently been cloned in the wrong orientation (Figure 3.17C), meaning that HR would result in integration of the plasmid backbone and the transgene would be lost (Figure 3.17D). Unfortunately, due to time limitations, I was unable to rectify this and reattempt the integration at hROSA26.

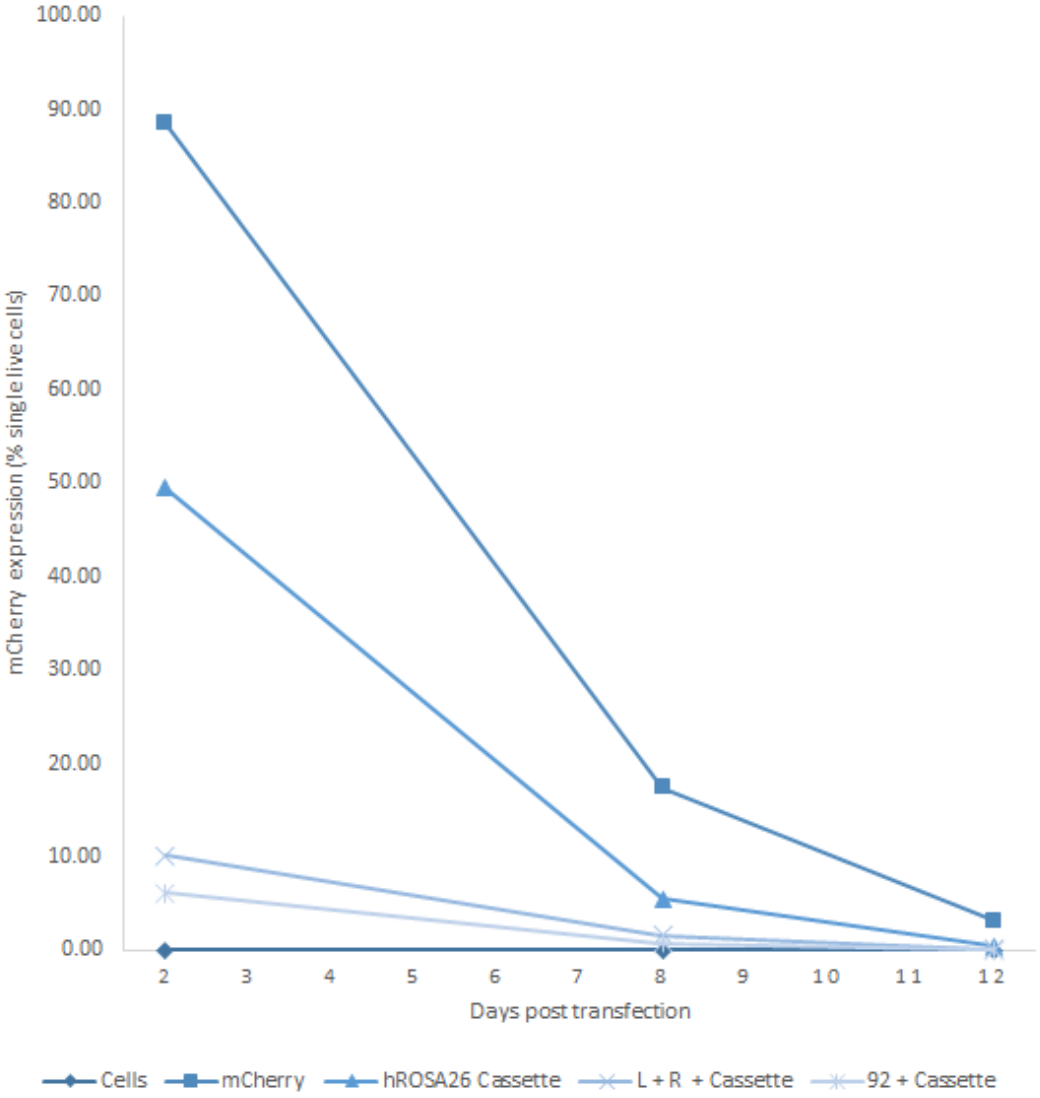


Figure 3.15 mCherry expression in heterogeneous cell populations. Following transfection, the mCherry expression in each experimental sample was monitored by flow cytometry for up to two weeks. The expression of mCherry rapidly declined in all experimental samples, with the majority showing <1% single live cells 12 days post transfection.

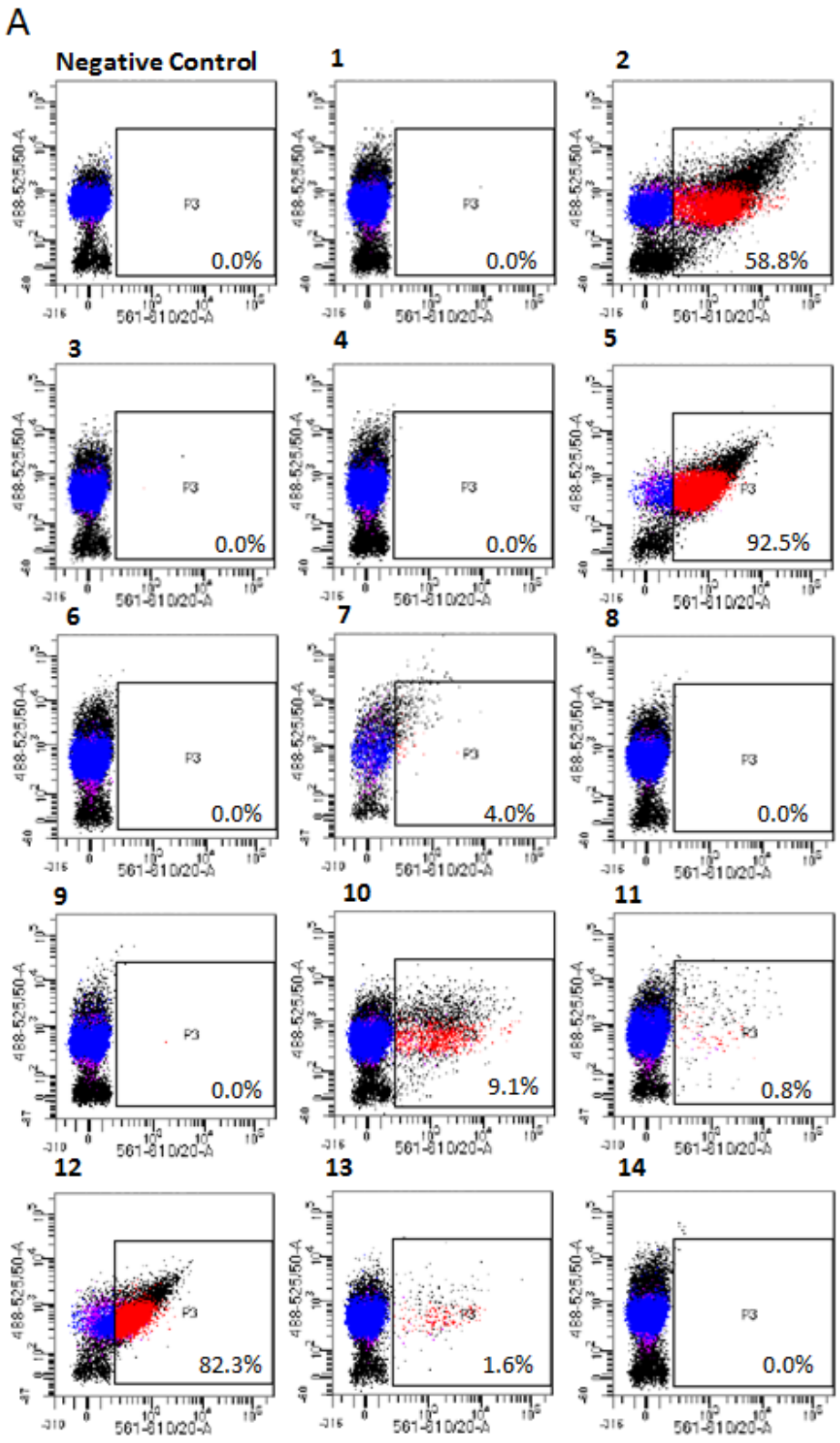
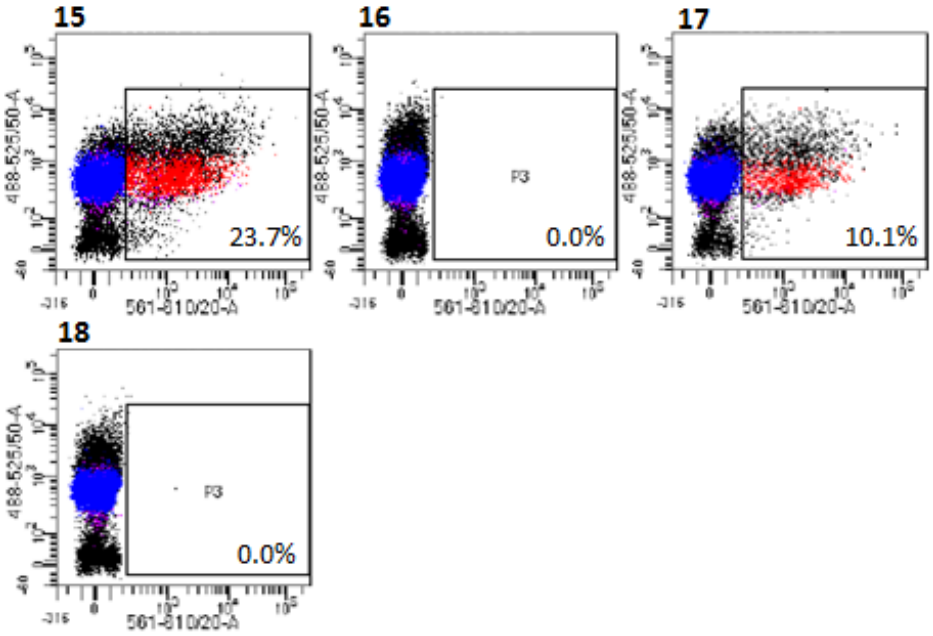


Figure 3.16 Integration at hROSA26. Page 1 of 2.



B

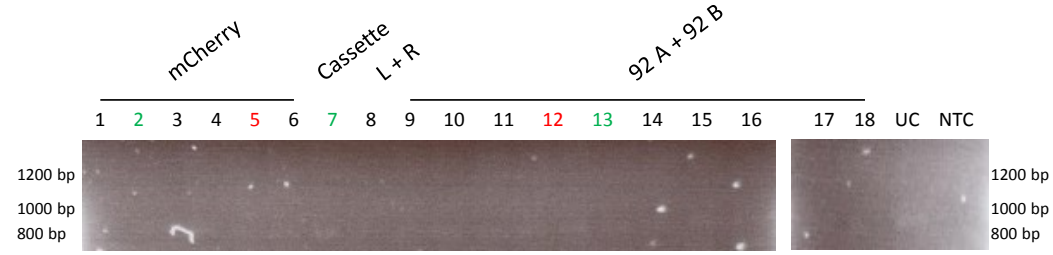


Figure 3.16 Continued. Page 2 of 2. A) Scatter plots of flow cytometric data for individual clonal lines. Clone number is indicated above each scatter plot. Scatter plots show mCherry expression levels indicated as percentage of live single cells. Red dots indicate mCherry positive, live single cells; blue dots are live cells; purple dots are live single cells; black dots are dead cells, or cell clumps. For each clonal line, 50,000 live single cells were analysed. B) PCR detection of correct integration for each clonal line. No lines demonstrated correct integration, indicated by the lack of a product band. Green numbers indicate mCherry expression between 1-80% live single cells. Red numbers indicate mCherry expression >80% live single cells. UC = untransfected control; NTC = no template control.

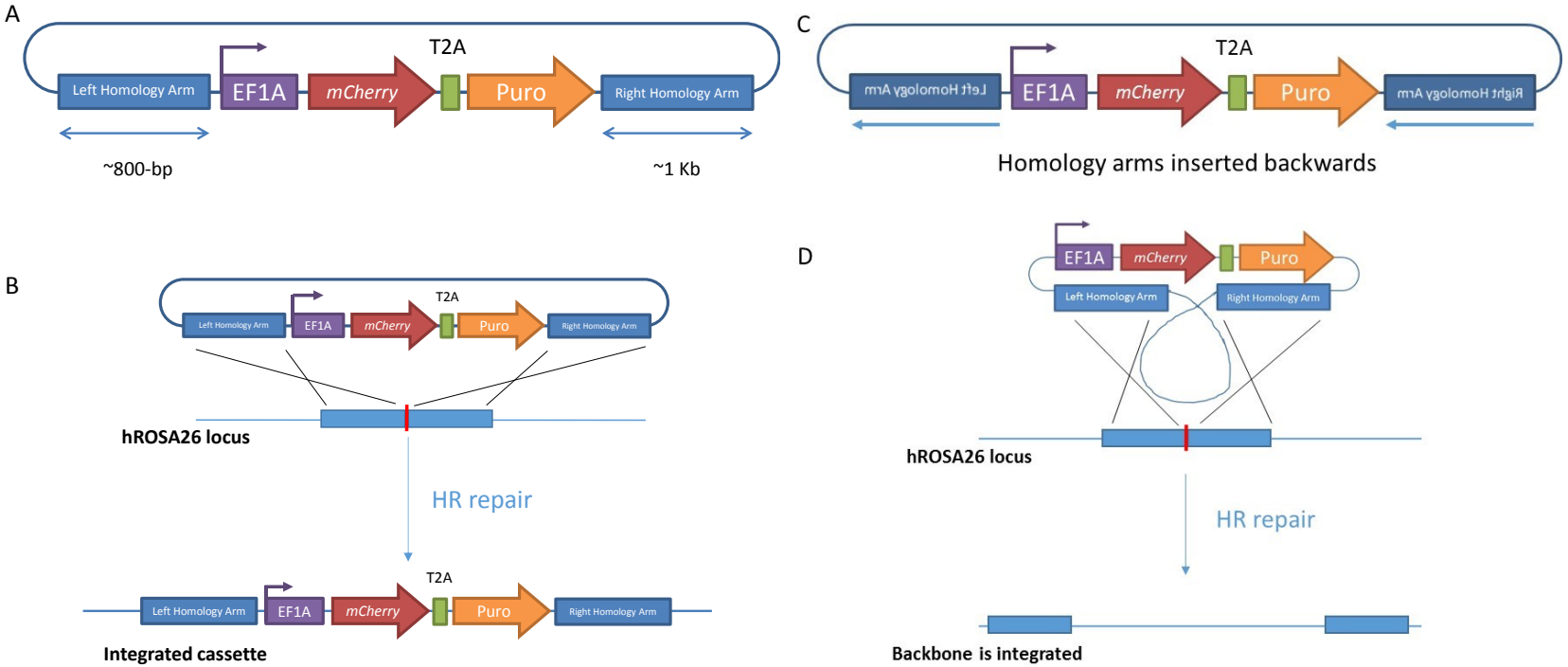


Figure 3.17 Integration at hROSA26. A) Diagram of hROSA26 targeting cassette. The cassette contains ~1.8-Kb total homology. B) Schematic of intended integration at hROSA26. Following cleavage by the hROSA26-specific SSNs, the DSB (red line) will be repaired via HR using the cassette as a donor template, resulting in the integration of the *mCherry* transgene unit. C) Diagram of hROSA26, demonstrating incorrect assembly of homology arms. Both homology arms were clones into the cassette in the reverse orientation, so correct integration of the mCherry transgene unit would not occur following HR repair. D) Schematic of predicted outcome of HR repair with the incorrectly assembled hROSA26 cassette. Due to the orientation of the homology arms, HR repair would lead to integration of the plasmid backbone, not the transgene unit.

3.4 Discussion

3.4.1 Target Regions are Transcriptionally Active

The qPCR data demonstrate that all the loci of interest are transcriptionally active within the panel of cell lines investigated. This is consistent with the ENCODE data for these regions, which indicate that they lie in areas of open chromatin. While these data imply that the chromatin is likely to be in an open confirmation at the target loci, they do not necessarily indicate the levels of expression that may be achieved upon transgene integration, as this will largely be dependent on the promoter within the targeting cassette. In addition, the presence of transcriptional activity of endogenous genes does not necessarily mean that silencing of transgenes integrated at these loci cannot occur. However, knowing that these areas are transcriptionally active suggests that these regions are in an environment that is more likely to be permissive to long-term gene expression.

3.4.2 Demonstrating Activity of SSNs

Production of the hROSA26 TALEN proteins *in vitro* yielded two products for the left TALEN, one of expected size and one of a lower molecular weight, approximately 73-KDa (Figure 3.2B). This could have been the result of protein degradation by phosphatases. This is unlikely as there is no evidence of this in any of the other samples, which were prepared simultaneously under the same conditions, and this is unlikely to result in a single distinct band. It may also have been the result of a cryptic promoter. This would have to be in the repeat region of the TALEN as the backbone is the same for all TALENs, thus if it were present in the backbone, one would expect to see additional bands for all the TALENs, which is not the case. However, this too is unlikely as extensive sequence analysis failed to find a start codon positioned to produce a fragment of the length observed (corrected for biotinylation). Therefore, a potential explanation is the presence of a cryptic splice site within the repeat region of the hROSA26 left TALEN, resulting in a smaller spliced fragment.

The results of the *in vitro* cleavage assay carried out with the hROSA26 TALEN pair suggest that the right TALEN may be able to act as a homodimer. The evidence for this is limited to the presence of one potential cleavage band following agarose gel

electrophoresis, which is a similar size to the larger of the two predicted cleavage products (Figure 3.2C). FokI nuclease activity is dependent on homodimerisation, meaning that each TALEN has the potential to homodimerise to bring about nuclease activity, resulting in a DSB, as long as appropriate binding sites are present; theoretically, the hROSA26 left TALEN can act as a homodimer. More recently, newer backbone plasmids have been developed which restrict activity to heterodimer formation due to mutations within the FokI nuclease domain (Naitou et al. 2015). Despite this, it seems unlikely based on sequence analysis that the right TALEN is acting as a homodimer in this instance. Indeed, under physiological conditions, no evidence of homodimer activity was observed (Figure 3.4A & B).

As previously stated, the ability of the AAVS1 TALEN pair to induce indels had been demonstrated by others (Sanjana et al. 2012). Initial attempts to replicate this were unsuccessful (Figure 3.2D), but I was subsequently able to confirm activity better than that observed previously (Sanjana et al. 2012). I was also able to confirm activity of gRNA_AAVS1_T2, observing an indel frequency of 21%, which is similar to that observed by Mali et al. (2013).

To demonstrate the utility of the CRISPR/Cas9 system, I also carried out the SURVEYOR assay. I was able to confirm the activity of all single sgRNAs in combination with wild type Cas9. However, in some cases activity is only apparent 72 hours post transfection. This suggests that indel frequency increases over time and that some less active sgRNAs require a longer period in which to accumulate indels above the detection threshold. This is consistent with the findings of Shalem et al. (2014b) who also showed that indel frequency accumulates over time.

The activities of the paired nickases seemed more variable than that of individual sgRNAs with wild type Cas9. When I designed the paired nickases for *IMPDH2* and *EEF1A1*, little was known about the effect of orientation on activity and no tools were available to identify optimal paired target sites, only single target sites. As such, I selected high scoring single target sites which occurred on opposite DNA strands within 30-bp of each other. Consequently, four out of the five pairs designed were in a head-to-head orientation; three of the four head-to-head oriented pairs – those targeting *IMPDH2* – did not show any observable activity despite each individual

sgRNA showing robust activity in the presence of wild type Cas9. However, both pairs targeting *EEF1A1* were able to demonstrate activity; one pair is in a head-to-head and the other is in a tail-to-tail orientation. Activity of paired nickases has been shown by others to be highly dependent on the orientation of the two sgRNAs with regards to each other; tail-to-tail oriented pairs tend to be more active than those oriented in a head-to-head configuration (Shen et al. 2014). This is consistent with my results for the *IMPDH2*-targeting pairs, which are oriented in a head-to-head configuration and fail to show activity in the SURVEYOR assay. However, the nickase pair targeting *EEF1A1* (Ex3) which are also in a head-to-head configuration show robust activity and do not appear to differ from the other nickase pair (Ex4) oriented tail-to-tail, which may indicate a context-dependent effect.

The nickase pair targeting hROSA26 was designed in light of these observations, and was designed in a tail-to-tail configuration to maximize chances of activity, (http://crispr.mit.edu/). Indeed, I was able to demonstrate its ability to induce indels at its target site (Figure 3.4B). However, in this case three – rather than the expected two – cleavage bands were present following digestion with SURVEYOR nuclease. Based on the approximate sizes of the cleavage products (~190-bp, ~130-bp, and ~100-bp) it would seem that the 61-bp overhang caused by the staggered DSB has been resolved in multiple ways. An indel close to sgRNA_hROSA26_92A would cause cleavage products of ~130-bp and ~160-bp, while an indel close to sgRNA_hROSA26_92B would cause cleavage products of ~100-bp and ~190-bp. Finally, if an indel were to occur between the two sites then cleavage products of ~130-bp and ~160-bp would be produced. Based on these predictions, it seems likely that the majority of indels have formed at the sgRNA__hROSA26_92B site. The presence of the 130-bp bands suggests that indels have formed either between the sites or at the sgRNA_hROSA26_92A site. In this case a band of ~160-bp should also be observed; this may be present but masked by the ~190-bp band, though a higher resolution agarose gel would be required to verify this. In addition, sequencing of individual amplicons would give a clearer idea of the range of indels following CRISPR/Cas9 activity at this locus.

During my investigation of SSN activity in other cell lines, I discovered that SNPs could complicate the interpretation of the SURVEYOR assay results (Figure 3.9). This has implications for the identification of sgRNA target sites; if SNPs are present at the target site, they also have the potential to impede the binding of the sgRNA. Consequently, it was not possible to investigate the activity of each SSN in different cell lines using the SURVEYOR assay. It has been noted that the activity of the CRISPR/Cas9 system can vary between different cell lines (Mali, Yang, et al. 2013). This may be due to differences in relative transfection efficiencies, or could suggest that cell line specific differences have a bearing on activity. These could either be in the relative activities of the HR and NHEJ pathways in different cells or due to epigenetic differences. Indeed, for TALENs, it has been demonstrated that CpG methylation can impede activity (Bultmann et al. 2012). Methods to overcome this have been proposed, which make use of the N* repeat, which is capable of binding to 5mC (Valton et al. 2012). On the other hand, 5mC has been reported to have little effect on the activity of CRISPR/Cas9 (Hsu et al. 2013; Perez-Pinera et al. 2013), which may be due to the tolerance of small mismatches between the sgRNA and the target site.

3.4.3 Transgene Integration

I was able to integrate an *EGFP* gene into the AAVS1 locus using both TALENs and CRISPR/Cas9, but not using the single nickase CRISPR/Cas9 system, using a circular plasmid repair cassette with ~1.6-Kb total homology. These are similar to the results obtained by Mali and colleagues (2013) who were able to use both TALENs and gRNA_AAVS1_T2 in combination with Cas9 to repair a modified locus by HR using a linear dsDNA repair template. However, they also demonstrated the ability of a gRNA_AAVS1_T2 in combination with Cas9n to drive HR. This may be due to differences in the repair templates (linear versus circular plasmid) or the nature of the HR event (repair of a locus by removal of a 68-bp fragment compared to integration of a transgene ~1.6-Kb). Interestingly, they showed that HR driven by CRISPR/Cas9 appeared to be faster than that driven by TALENs, suggesting that TALEN-driven HR can take up to 40 hours. This is not consistent with the results I obtained; both systems resulted in transgene integration after 24 hours; however, this may be due to the differences in the detection of successful HR

events. In addition, they were able to demonstrate transgene integration using wild type Cas9 at the endogenous AAVS1 locus using both ssDNA oligonucleotides and dsDNA donor templates, consistent with the results here.

Building on this, I wanted to demonstrate that hROSA26 could also support transgene integration, driven by both TALENs and CRISPR/Cas9, and hoped to compare the two systems at this locus. As the AAVS1 region has been more extensively studied using ZFN (Chang & Bouhassira 2012; Wang et al. 2012; Zou et al. 2011; DeKelver et al. 2010; Smith et al. 2008), TALENs (Cerbini et al. 2015; Mussolino et al. 2014; S. Zhang et al. 2014; Maggio et al. 2014; Hockemeyer et al. 2011) and CRISPR/Cas9 (Cho et al. 2014; Maggio et al. 2014; J.-H. Zhang et al. 2014; Mali, Yang, et al. 2013), in addition to the appearance of commercially available integration systems based upon AAVS1 (Sigma-Aldrich® Cat. No. CTI1; GeneCopoeia Cat. No. SH-AVS-K100), I decided to focus my efforts on further investigating hROSA26. To achieve this, I constructed a targeting cassette containing an *mCherry* gene and a puromycin resistance cassette under the control of the EF1α promoter, with approximately 1.8-Kb total homology. Following transfection and selection with puromycin, there was a large difference in the proportions of mCherry positive cells in each population, even though each population was transfected with the same amount of mCherry-expressing plasmid, with the TALENs and CRISPR/Cas9 populations having the lowest levels of mCherry expression (Figure 3.20). Subsequently the proportion of mCherry expressing cells in each population rapidly declined. After sorting for mCherry-expressing cells, the proportions of mCherry positive cells in each population remained low (less than 10% single live cells). Due to the low numbers of mCherry positive cells, I was only able to generate 18 clonal lines. Of these, only 2 lines showed high level mCherry expression (greater than 80% live single cells) and none demonstrated correct insertion of the cassette at the target locus (Figure 3.16A & B). This was unexpected based on the results from the AAVS1 locus. However, following further inspection, it became apparent that the homology arms within the targeting cassette had been inadvertently cloned in the wrong orientation (Figure 3.17C) so that integration of the plasmid backbone and not the transgene would occur following HDR of the SSN-induced DSB (Figure 3.17D). This explains the lack of correct integration at the target locus and helps to explain

the continued loss of mCherry expression even after enrichment by FACS; upon integration, the mCherry gene would be lost. It is likely that the clones showing high-level mCherry expression result from the random integration of the transgene and enrichment by FACS has selected for those integrated into loci that permit robust expression. It would be worth investigating these clones further and identify the integration loci, which may be useful for further studies of transgene integration.

Some clones demonstrated low-level mCherry expression. These clones are likely to harbour random integration of the transgene. However, in these cases, integration loci are less permissible to robust expression, or have been subject to incomplete silencing. The variegated phenotype observed, is likely due to position effects, which are commonly seen following random integration.

3.5 Future Perspectives

Due to technical issues and time constraints, I was unable to demonstrate successful transgene integration at the hROSA26 locus. Therefore, I could not investigate long-term expression of transgenes at this locus. It remains important to characterise this site as a suitable locus for prolonged expression. The mouse *Rosa26* site has been used successfully for transgene expression since its identification over two decades ago (Friedrich & Soriano 1991), which suggests that the human homologue may be an ideal location for efficient transgene expression. Indeed, Irion et al. (2007) demonstrated the ability of hROSA26 to support transgene expression even after embryonic stem cell differentiation into a range of lineages.

More recently, homologues of the *Rosa26* locus have been identified in rat (Kobayashi et al. 2012) and pig (Li et al. 2014), which can both support robust transgene expression. This suggests there has been evolutionary conservation of this locus; it would be worth investigating how well conserved it is across other mammalian and non-mammalian species and if all can support transgene integration.

Careful choice of transgene integration locus can reduce the risk of subsequent transgene silencing, but other factors can also influence this, particularly the design of the repair cassette. Studies have shown that different promoters show individual susceptibilities to silencing (Spencer et al. 2015; Gray et al. 2011; Papadakis et al.

2004). Judicious promoter choice is likely to be necessary for efficient and prolonged expression, though identification of appropriate promoters for each target locus may only be possible through experimentation. In addition, orientation of integration may affect the expression levels. Chen et al. (2011) showed that certain promoters showed orientation dependent activity at the *Rosa26* locus. Additional cassette design considerations can be made in order to maximize the likelihood of optimal transgene expression such as inclusion of insulators, though each would likely need to be validated experimentally in order to find the optimal cassette design.

One of the potential outcomes of this work was the development of a stable producer cell line for recombinant lentiviral vectors, currently being developed within the UKCFGTC. Recently, Sanber and colleagues (2015) have generated a stable producer line in which they identified suitable sites for expression by screening clones generated following integration of gammaretroviral vectors. They used a non-cytotoxic envelope protein (RDpro, derived from gammaretroviral RD114 envelope glycoprotein), but do suggest that inducible promoters could be used should a cytotoxic envelope protein be required.

This system demonstrates the feasibility of a producer line, but would need to be reengineered for each application (e.g. different transgene of interest). To achieve this it would be advantageous to have sites known to support expression with an SSN toolkit to facilitate integration. This has been demonstrated for AAVS1, and with more investigation could prove to be the case for hROSA26. The tools I developed here targeting hROSA26, along with the SSNs targeting AAVS1, will be useful for this purpose.

Chapter 4 Development of a Drug-Selectable Transgene Integration System

4.1 Introduction

Transgene integration is a vital tool in cell biology. However, current methods for integrating transgenes can be inefficient and usually rely on selection for drug resistance genes forming part of the targeting cassette. Transgenes can integrate either site-specifically or randomly into the genome of their recipient cell, driven respectively by HR and NHEJ DNA-repair pathways (Sakurai et al. 2010). Other pathways, such as the alternative end-joining pathway (Iiizumi et al. 2008) and microhomology-dependent random integration (Merrihew et al. 1996), have also been implicated in the random integration of transgenes. Site-specific transgene integration requires targeting cassettes carrying large regions of homology to the intended target locus, which can be difficult to construct. In mammals this occurs less commonly than random integration (Zheng & Wilson 1990) as NHEJ occurs more frequently than HR (Sakurai et al. 2010). However, randomly integrated transgenes are prone to position effects and gene silencing, resulting in unreliable and unpredictable expression (Sadelain et al. 2012).

Newer technologies, particularly the advent of SSNs, have helped to overcome some of these issues; transgenes can be efficiently targeted to predetermined loci, selected for their ability to permit robust and long-term expression, which would reduce the risk of gene silencing. In addition, the production of a DSB at the target site increases the HR efficiency in mammalian cells by several orders of magnitude (Elliott et al. 1998; Liang et al. 1996; Choulika et al. 1995; Rouet et al. 1994a). However, selecting for site-specific integration events remains reliant on the introduction of a selectable marker.

To help address this issue, I designed a novel drug-selectable transgene integration system. The intent was to allow rapid selection of transgenes integrated at the target locus (*IMPDH2*), whilst eliminating cells containing randomly integrated transgenes, without the need for an exogenous selection marker.

4.1.1 Drug-Selectable Transgene Integration System

The principle of the proposed selection system is based on the fact that the activity of both human inosine monophosphate dehydrogenase (IMPDH) isoforms can be inhibited by the drug mycophenolic acid (MPA), while variants of IMPDH are

known to be severalfold more resistant (discussed in Section 4.1.4). The *IMPDH2* locus will be targeted using the CRISPR/Cas9 system, resulting in a DSB (Figure 4.1). This will subsequently be repaired via HR using the repair cassette provided. The repair cassette will consist of the distal portion of an *IMPDH* cDNA variant (which encodes an IMPDH variant severalfold more resistant to MPA than IMPDH2) and a transgene (initially a reporter gene, such as *mCherry*) in a separate transcriptional unit flanked by arms of homology. Once repaired, the locus will encode a hybrid IMPDH enzyme capable of conferring resistance to MPA. IMPDH is a key metabolic enzyme (Section 4.1.2); its inhibition by MPA should be lethal, thus only cells bearing the correct insertion will survive selection in MPA-containing media (Figure 4.1). This should result in robust long term expression of the transgene which can be observed through monitoring the mCherry signal over time.

To ensure that the repair cassette encodes a functional enzyme, the CRISPR/Cas9 system will be used to generate a DBS within intron five of the endogenous *IMPDH2* locus, and the IMPDH variant will contain the endogenous splice acceptor site from *IMPDH2* to ensure correct processing. The *IMPDH2* locus is organised into 14 exons and 13 introns spanning 5118-bp on human chromosome 3 (position 3p21.31). The enzyme's active site is encoded by exons nine and ten, and the residues which make direct contact with MPA are encoded by exons eight, nine and twelve. For the purpose of the drug-selectable transgene integration system, I needed to target the *IMPDH2* locus upstream of these features, in order to generate a functioning hybrid which could confer MPA-resistance to cells. I therefore chose to target intron five. It is the largest intron within the *IMPDH2* locus and so would give greatest flexibility when designing sgRNAs. In addition, it is distal enough from the transcription start site that the repair cDNA should not of itself encode an active enzyme, so if random integration should occur, it is unlikely that MPA resistance would be conferred.

Unlike other currently available selection systems, such as those relying on exogenous selection markers within the repair cassette, the attractive feature of this system lies in the fact that only transgenes correctly integrated at the *IMPDH2* locus will result in a functional IMPDH variant; the partial cDNA encoding the IMPDH variant within the transgene cassette will not produce a functional enzyme unless

correctly integrated at the *IMPDH2* locus, thus only cells with a correctly integrated transgene will be resistant to MPA.

4.1.2 Inosine Monophosphate Dehydrogenase

IMPDH (EC 1.1.1.205) is the enzyme responsible for the oxidation of inosine monophosphate (IMP) to xanthosine monophosphate (XMP), with concomitant reduction of nicotinamide adenine dinucleotide (NAD), the first committed and rate limiting step in *de novo* guanine nucleotide biosynthesis (Figure 4.2). Humans have two isoforms, IMPDH1 and IMPDH2 (Natsumeda et al. 1990), which are encoded by the genes *IMPDH1* and *IMPDH2,* respectively. These isoforms are 84% identical at the amino acid level and have almost indistinguishable kinetic properties. Both isoforms are expressed to varying degrees in most tissues, though IMPDH1 is more highly expressed in the retina, spleen and resting peripheral blood mononuclear cells (Hedstrom 2009), while IMPDH2 is upregulated during proliferation and in some cancers (Hedstrom 2009). In addition, IMPDH1 null mice present with retinal degeneration (Aherne et al. 2004) while IMPDH2 knockouts die in early embryogenesis (Gu et al. 2003).

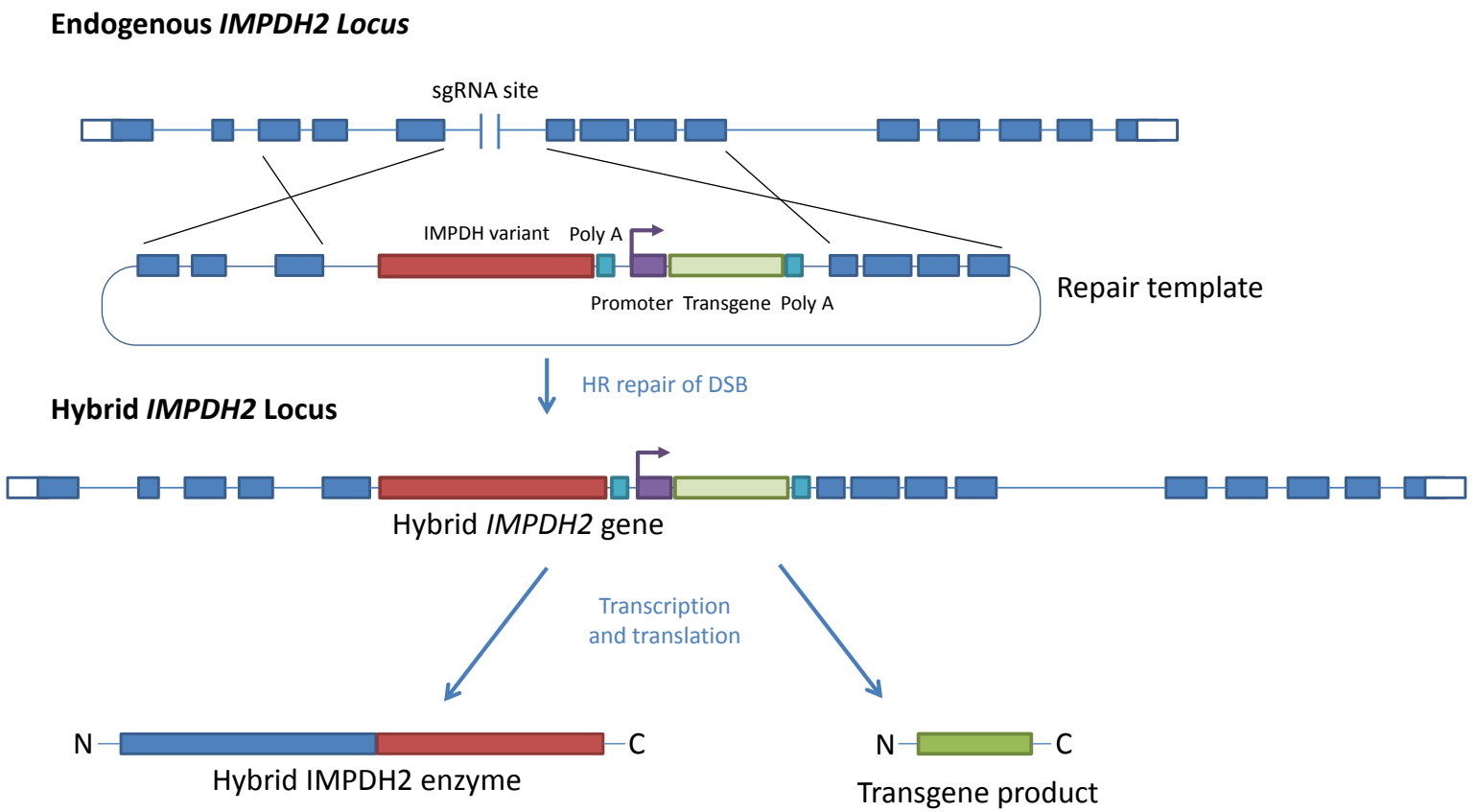


Figure 4.1 Proposed drug-selectable transgene integration system at *IMPDH2*. The human *IMPDH2* locus will be targeted using the CRISPR/Cas9 system, resulting in a DSB. This will be repaired by HR using the repair cassette provided. The repair cassette will consist of the distal portion of an *IMPDH* variant along with a transgene of interest (initially *mCherry*) flanked by arms of homology. Once integrated, the IMPDH variant will form a hybrid enzyme with the proximal portion of IMPDH2, which will confer cells with an increased resistance to MPA, allowing for rapid selection of correctly integrated transgenes.

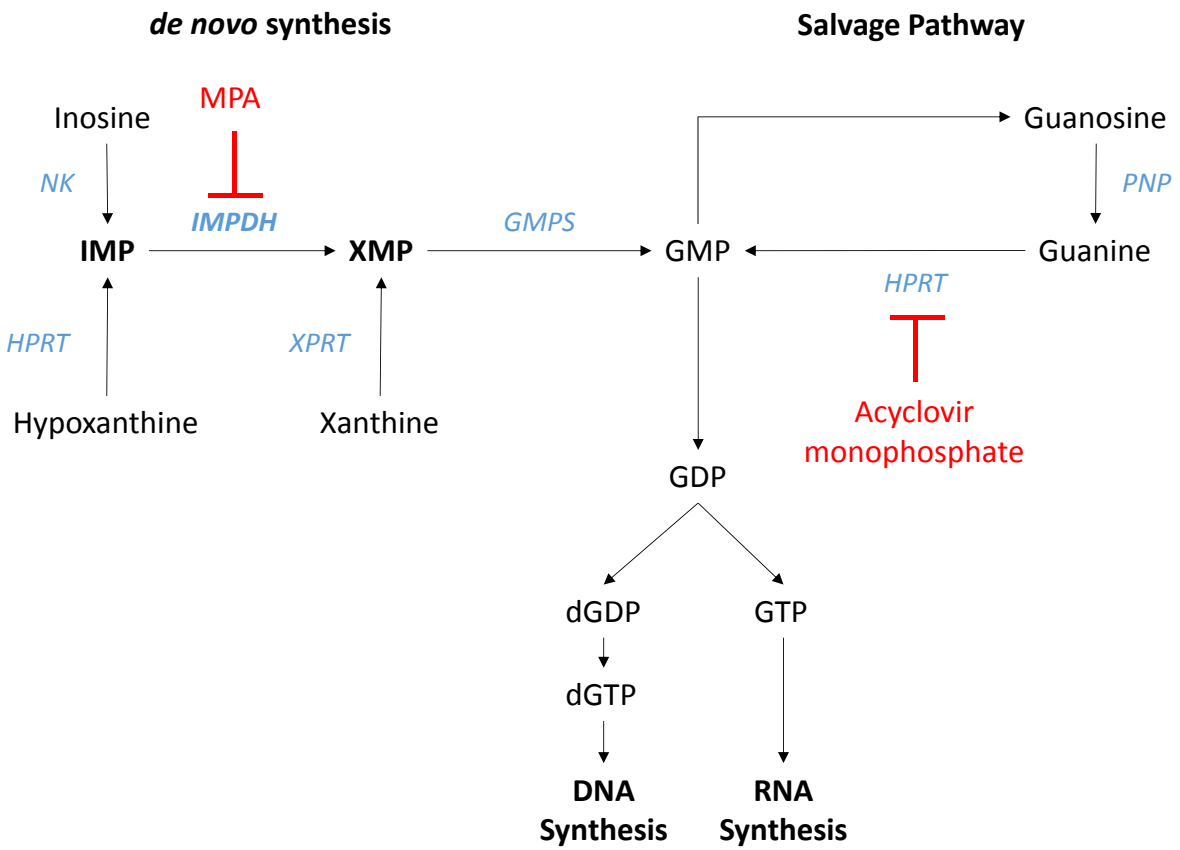


Figure 4.2 Guanine nucleotide biosynthesis. IMPDH catalyses the conversion of IMP to XMP (bold) which is the first committed and rate limiting step in guanine nucleotide biosynthesis. The salvage pathway recycles guanosine and guanine into GMP which can then be used for DNA and RNA synthesis. Key enzymes are indicated in blue. Inhibitors are depicted in red. IMP = inosine monophosphate; XMP = xanthosine monophosphate; GMP = guanosine monophosphate; GDP = guanosine diphosphate; GTP = guanosine triphosphate; dGDP = deoxyguanosine diphosphate; dGTP = deoxyguanosine triphosphate; NK = nucleoside kinase; HPRT = hypoxanthine-guanine phosphoribosyl transferase; XPRT = xanthine phosphoribosyl transferase; GMPS = GMP synthase; PNP = purine nucleoside phosphorylase; MPA = mycophenolic acid.

4.1.2.1 Guanine Nucleotide Synthesis

Cells have two pathways for delivering their guanine nucleotide needs: the *de novo* synthesis pathway and the guanine salvage pathway. IMPDH is essential for the *de novo* synthesis pathway, catalysing the first committed and rate limiting step (Figure 4.2). In addition to this pathway, cells can also recycle guanine via the salvage pathway (Figure 4.2). The main enzyme responsible for this is hypoxanthine-guanine phosphoribosyl transferase (HPRT). The activity of these pathways varies depending on cell type, with rapidly proliferating cells such as lymphocytes relying almost entirely on the *de novo* synthesis pathway for their guanine nucleotide needs (Allison et al. 1977).

4.1.2.2 Role of IMPDH in Disease

IMPDH is a key regulator of the guanylate pool, which is essential for many metabolic pathways and as such its activity is closely linked with oncogenesis. Indeed, IMPDH, primarily IMPDH2, activity and expression is significantly elevated in many human and animal tumours (as reviewed by Oláh et al. 2006). In addition, it has been demonstrated that IMPDH1 plays a key role in angiogenesis (Chong et al. 2006; Wu et al. 2006).

T and B lymphocytes, important modulators of the immune system, are reliant on the *de novo* synthesis pathway for guanine nucleotides and proliferation. Thus, inhibitors of IMPDH can be utilised for immunosuppression (Jain et al. 2001); for example, the morpholinoethyl ester of MPA, mycophenolate mofetil (MMF), is commonly used as an immunosuppressive agent to prevent rejection following solid organ transplant (Neyts et al. 1998).

In addition, microbial IMPDH enzymes are sufficiently distinct from the human forms; in recent years, there has been an increasing interest in the development of specific IMPDH inhibitors for use in antimicrobial therapy (Braun-Sand & Peetz 2010; Nair & Shu 2007).

4.1.3 Mycophenolic Acid

MPA is a natural compound first isolated from the fungus *Penicillium brevicompactum* and forms the active metabolite of the widely used

immunosuppressant MMF. MPA is a specific and potent inhibitor of mammalian IMPDH, inhibiting both human isoforms, though it is much less potent for microbial IMPDH enzymes. It inhibits IMPDH by trapping the IMP reaction intermediate (XMP*) in the enzyme's active site (E-XMP*), rendering it inactive (Fleming et al. 1996; Link & Straub 1996). Inhibition of IMPDH results in a depletion of the guanylate pool, particularly GMP, and inhibition of DNA synthesis, leading to cessation of cell proliferation.

4.1.4 MPA-Resistant IMPDH Variants

For the purpose of this study, I am interested in two IMPDH variants which have been shown to have increased resistance to MPA compared to human IMPDH2: *Tritrichomonas foetus* IMPDH (*Tf*-IMPDH) and IMPDH^{IY}. *Tf*-IMPDH from the protozoan *T. foetus* is 500-fold more resistant to MPA than human IMPDH2 (Hedstrom 2009). Studies have shown that while some of this resistance is due to changes in the residues directly contacting MPA, not all of the resistance can be attributed to this (Digits & Hedstrom 1999). Thus, this may add additional complexity to the design of the drug-selectable integration system. IMPDH^{IY} is a variant of IMPDH first identified in neuroblastoma cells selected for 10,000-fold increased resistance to MPA treatment (Hodges et al. 1989). It carries two amino acid substitutions: T333I and S351Y. Thr333 is one of the amino acids which directly interact with MPA: T333I alone has been shown to increase the K_i of MPA 300-fold (Sintchak et al. 1996). IMPDH^{IY} contains only two amino acid substitutions compared with wild type IMPDH2, and would thus make the design of the repair cassette more straightforward.

4.1.4.1 Comparisons with HPRT and the HAT Selection System

At first glance, the drug-selectable transgene integration system proposed here may seem similar to the HAT (hypoxanthine-aminopterin-thymidine) selection system based on the *HPRT1* gene (which encodes HPRT). However, a few important differences exist. Firstly, cells deficient in HPRT are resistant to the cytotoxic effects of 6-thioguanine (6-TG). Therefore, one could imagine integrating transgenes at the *HPRT1* locus, causing loss of HPRT function, then selecting for integration with 6-TG. However, this would also select for cells which carry loss of function *HPRT1*

mutations, due to NHEJ repair of the targeted DSB, but which do not have an integrated transgene. As NHEJ occurs more frequently than HR (Sakurai et al. 2010), one could postulate that the majority of selected clones would carry loss of function mutations and not integrated transgenes. Thus, selection would not be specific.

In contrast, a cell line null for HPRT and therefore sensitive to HAT selection could be used. In this instance, the targeting cassette would contain the *HPRT1* gene and thus integration would render cells resistant to HAT. However, selection of integration events would not necessarily be site-specific, as *HPRT1* integrated anywhere would confer resistance to HAT. Alternatively, a partial *HPRT1* cDNA could be used to repair the loss of function mutations at the *HPRT1* locus, analogous to the novel system proposed here. However, both of these methods rely on the availability of an HPRT null cell line; this would usually require time-consuming preliminary work to generate *HPRT1* loss of function mutations in a cell line of choice.

Another unique aspect to the system proposed here is that once validated, it will allow for the integration and selection of any transgene of interest, without the need for constructing complex targeting constructs with exogenous selection genes. Changing the transgene of interest will involve a single subcloning step to swap the initial reporter gene for any gene of interest. This can then be targeted to *IMPDH2* and correct integration selected for by MPA treatment.

4.2 Aims

The overall aim described in this chapter was to develop a drug-selectable transgene integration system at the human *IMPDH2* locus. This can be subdivided into the following in order to achieve a functional drug selection system:

1. Design and construct sgRNAs to target intron five of *IMPDH2* and assess their activity *in cellulo.*
2. Assess the ability to select IMPDH variant using MPA treatment.
3. Demonstrate the ability of *IMPDH2* to support transgene expression and robust long-term expression (greater than one month post transfection).

4.3 Results

4.3.1 Targeting *IMPDH2* with sgRNAs

I used the CRISPR/Cas9 system rather than TALENs for the development of this system as they have been shown to be equally effective, but sgRNAs can be more easily designed and constructed compared with TALENs (as discussed in Chapter 3).

4.3.1.1 Identifying sgRNA Target Sites

I identified potential sgRNA target sites using the CRISPR design tool developed by the Zhang laboratory (Hsu et al., 2013; available at http://crispr.mit.edu/; described in Chapter 3). This gave 22 possible sgRNA target sites. From these I selected six of the highest scoring sites which could be used to make three nickase pairs, binding opposite DNA strands with less than 30-bp between their target sites. Nickase pairs have a reduced off-target profile, thus should help to make the system as specific as possible. The genomic location of these targets is depicted in Figure 4.3.

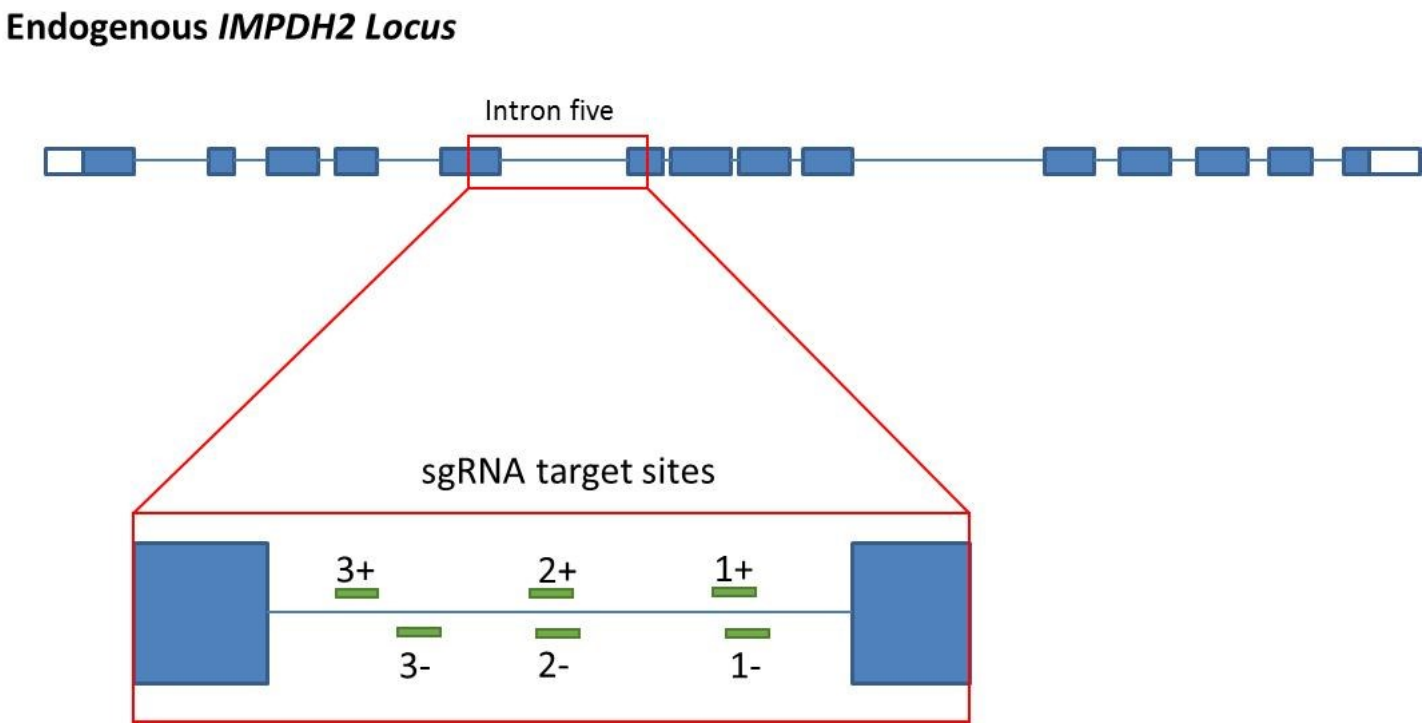


Figure 4.3 Location of *IMPDH2*-targeting sgRNAs on human chromosome 3. Green boxes indicate the location of the target sites for each individual sgRNA within intron 5 of *IMPDH2*. (+) denotes a target on the sense strand and (–) denotes a target on the antisense strand.

4.3.1.2 Validation of sgRNA Activity *in cellulo*

The activity of each individual sgRNA (in combination with wild type Cas9), as well as each nickase pair (in combination with Cas9n), was determined using the SURVEYOR assay and is described in detail in Chapter 3. In summary, each individual sgRNA was able to target Cas9 cleavage to the intended target locus, but all three nickase pairs failed to demonstrate detectable activity.

In order to confirm the results of the SURVEYOR assay and observe the range of indels generated following CRISPR/Cas9 activity, I generated clonal lines following transfection with each sgRNA and wild type Cas9 or each nickase pair by FACS. I isolated a total of 11 clonal lines derived from wild type Cas9 and 5 clonal lines from Cas9n. Unfortunately, some lines were lost during this process so I was not able to isolate clonal lines for all of the individual sgRNA. Following PCR amplification of the target region, I cloned and sequenced individual amplicons, the results of which are depicted in Figure 4.4.

Target efficiency is high with 10/11 (90.9%) clonal lines derived from wild type Cas9 having an indel at the target site. The observed mutations generally clustered around the predicted cleavage site (3-bp upstream of the PAM), though not always. In the case of sgRNA_3+, both insertion and deletion occurred downstream of the predicted cleavage site (16-bp and 13-bp, respectively). The most commonly observed mutations were deletions, ranging from 2-bp to 27-bp. Insertion events seem to be restricted to single bases, though in one case there was a large insertion of 237-bp, which aligns to a region in human chromosome 20 that shares 4-bp microhomology with the target. Some clones had more than two mutant alleles. Clones derived following transfection with the nickase pairs had no observable indels at the target sites, consistent with the data from the SURVEYOR assay.

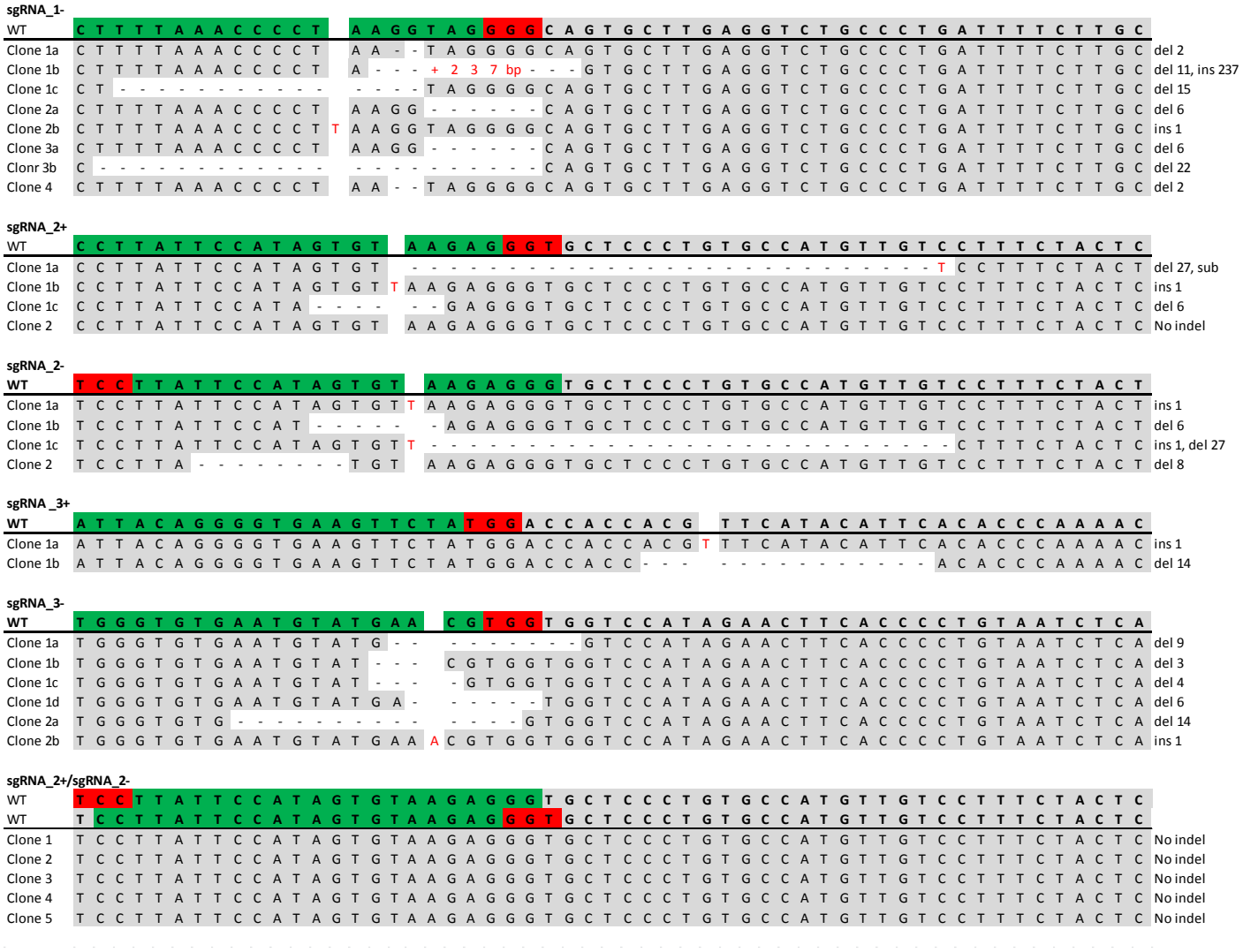


Figure 4.4 Range of mutations observed following CRISPR/Cas9 activity at the *IMPDH2* locus. Green boxes = sgRNA target; Red boxes = PAM; Grey boxes = no change; white box = indel; (-) = deletion; red letter = insertion; del = deletion; ins = insertion; WT = wild type.

4.3.2 Validation of MPA Sensitivity

Due to the gene therapy interests of my group, I decided to use HEK293FT cells for the creation of the drug-selectable integration system. This cell line is commonly used in the production of clinical grade lentiviral vectors and is a highly transfectable derivative of the HEK293 cell line, which possesses a stable integration of the SV40 large T antigen, required for efficient lentiviral production (Gama-Norton et al. 2011). In order for efficient selection of MPA-resistant IMPDH variants, the host cell line (HEK293FT cells) must show sensitivity to MPA treatment. Qasim et al. (2011) showed that HEK293 cells are sensitive to treatment with MPA (Figure 4.5B). In order to assess the sensitivity of HEK293FT cells to MPA treatment and determine any difference in response to HEK293 cells, I carried out a BrdU-based cell proliferation enzyme-linked immunosorbent assay (ELISA) following treatment with either 10 μM or 100 μM MPA for a period of 72 hours. This involved incubating cells with BrdU labelling agent for a period of 2 hours before fixing, permeabilising and staining with an anti-BrdU antibody conjugated to peroxidise. Total BrdU incorporation (as a marker of cell proliferation) was determined by colorimetric readout of the peroxidise-substrate reaction (measured using a plate reader).

Figure 4.5A demonstrates the dose-dependent sensitivity of both HEK293 cells and HEK293FT cells to MPA, with both 10 μM and 100 μM MPA, giving reduced levels of cell proliferation compared with the control (vehicle only). There is no significant difference between either cell line at both MPA concentrations tested ($p = 0.62$ for 10 μM and $p = 0.64$ for 100 μM). Although these results do not fully replicate the findings of Qasim et al. (2011; Figure 4.5B), who were able to show proliferation levels of less than 10% of control for the highest MPA concentration tested (100 μM), they do demonstrate sensitivity of both these cell lines to short term (72 hour) treatment with MPA; this ought to have been sufficient to allow for effective selection of MPA resistant variants of IMPDH.

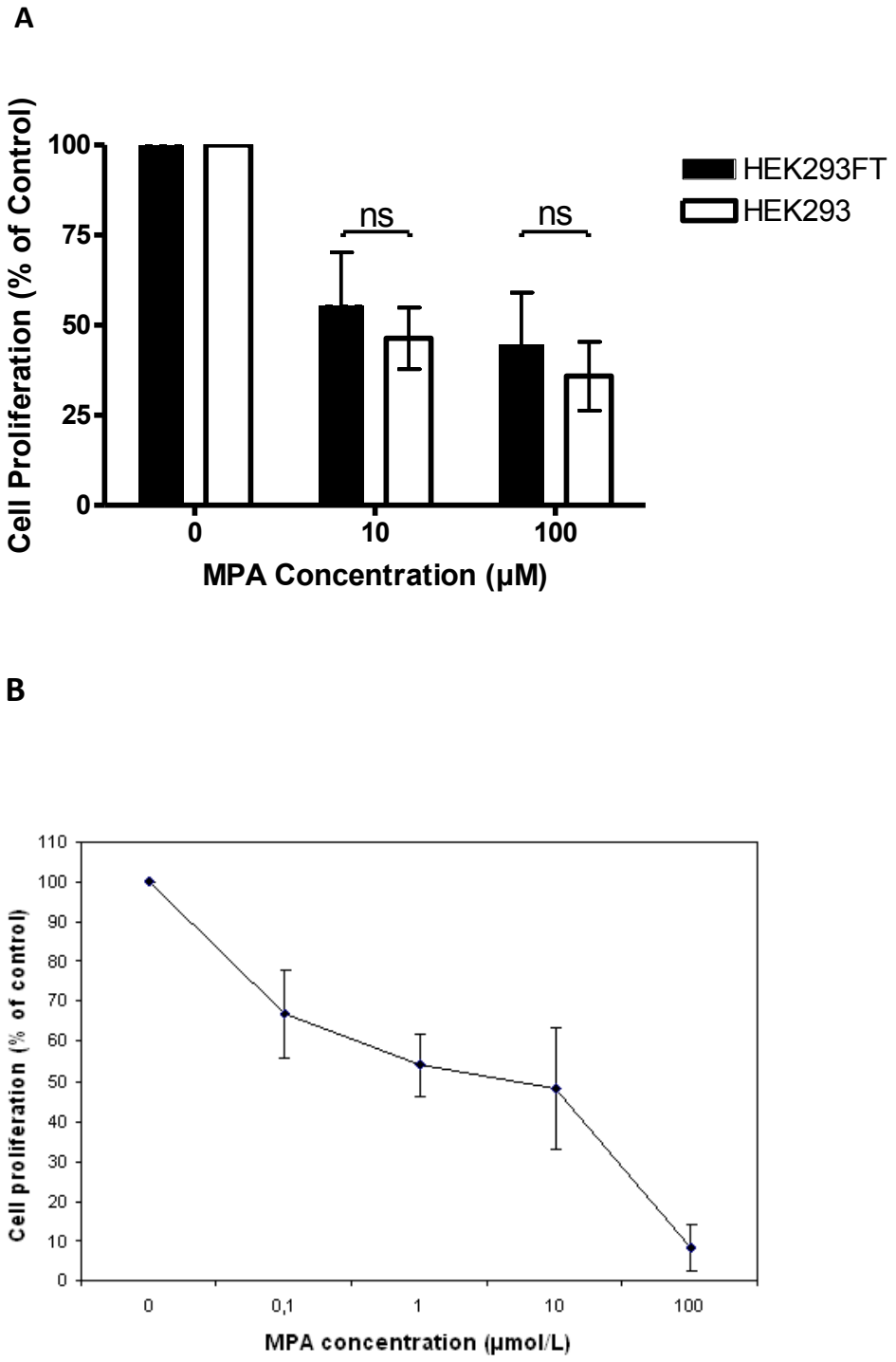


Figure 4.5 HEK293 and HEK293FT cell survival following MPA treatment. A) Cells were treated with 0, 10 or 100 µM MPA for a period of 72 hours. Cell proliferation was determined using a BrdU based ELISA. Results are shown as percentage of untreated control (0 µM MPA). No significant difference was found between HEK293 and HEK293FT at either 10 µM or 100 µM MPA. Means of 3 independent experiments are shown. Error bars denote standard error of the mean (S.E.M.). ns = not significant ($p > 0.05$). B) Figure taken with permission from Qasimet al. (2011) showing the effect of a range of MPA concentrations on HEK293 cell proliferation, determined using a BrdU-based ELISA.

4.3.3 Validation of IMPDH2 Variants

To assess the ability of IMPDH^{IY} and *Tf*-IMPDH (discussed in Section 4.1.4) to confer increased resistance to MPA, compared with wild type IMPDH2, I carried out the same BrdU-based ELISA described above. In this instance, cells cultured in 96-well plates were first transfected with cDNA expression vectors encoding each variant (or a wild type IMPDH2, to control for changes in resistance due to overexpression). Transfection efficiency was estimated using a concurrent transfection with an EGFP expression vector, with fraction of EGFP positive cells determined using a fluorescence microscope. Transfection efficiency was ≥60%. Cells were then treated with either 10 μM or 100 μM MPA for a period of 72 hours, 24 hours post transfection.

At the lowest MPA concentration tested (10 μM), neither IMPDH^{IY} nor *Tf*-IMPDH demonstrated greater resistance than overexpression of IMPDH2, with none of the three reaching significance over wild type (untransfected) cells (Figure 4.6). Interestingly, IMPDH^{IY} did not confer increased resistance to MPA treatment even at the highest concentration tested (100 μM) compared to wild type cells and performed worse than overexpression of IMPDH2. Thus, IMPDH^{IY} was excluded from further experiments. However, *Tf*-IMPDH did confer increased resistance to MPA compared to wild type at both MPA concentrations tested, as demonstrated by the higher level of proliferation; this only reaches significance at 100 μM MPA ($p = 0.0066$), where it also demonstrates significantly more resistance to MPA than IMPDH2 overexpression and IMPDH^{IY}($p = 0.0242$ and $p = 0.0053$, respectively).

While *Tf*-IMPDH was able to confer significantly elevated resistance to MPA treatment compared to wild type (and IMPDH2 overexpression) at 100 μM, there was still a high proportion of wild type cells surviving the MPA treatment (approximately 60% of untreated control), making selection of MPA resistant cells difficult. It is clear that *Tf*-IMPDH demonstrates increased resistance to MPA treatment over wild type cells, but this difference between *Tf*-IMPDH-expressing cells and wild type cells, though statistically significant, would not underpin a practical selection system. To address this, I next tested whether extending the treatment time would improve the selection of MPA-resistant IMPDH variants. I

therefore carried out the BrdU-based cell proliferation ELISA after a period of 120 hours of MPA treatment on cells transfected with either *Tf*-IMPDH or IMPDH2.

Tf-IMPDH does not confer significant resistance to MPA compared to wild type HEK293FT cells or cells overexpressing IMPDH2 at 10 μM MPA following either 72 hours or 120 hours of treatment (Figure 4.7). However, at 100 μM MPA, it confers significant resistance following both 72 hours and 120 hours of treatment over wild type and IMPDH2 overexpression. While *Tf*-IMPDH confers significant resistance at both time points, residual survival of wild type cells is reduced following 120 hours of treatment; I used 120 hour treatment for all subsequent experiments.

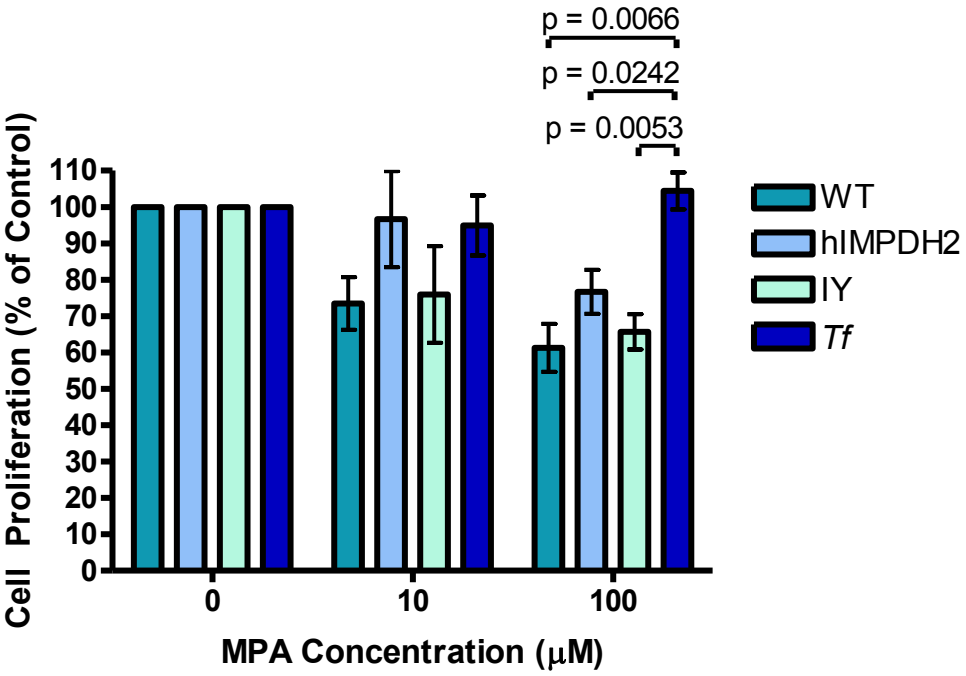


Figure 4.6 Validation of MPA resistance of IMPDH2 variants. Cells were transfected with either human IMPDH2 (hIMPDH2), human IMPDH2^{IY} (IY) or *Tf*-IMPDH (*Tf*) cDNA expression vectors prior to treatment with 0, 10 or 100 µM MPA for a period of 72 hours. Cell proliferation was determined using a BrdU-based ELISA. Results are shown as percentage untreated control (0 µM MPA). Data represent means of 3 independent experiments. Error bars show the standard error of the mean (S.E.M.). Unpaired T-tests were used to calculate p-values.

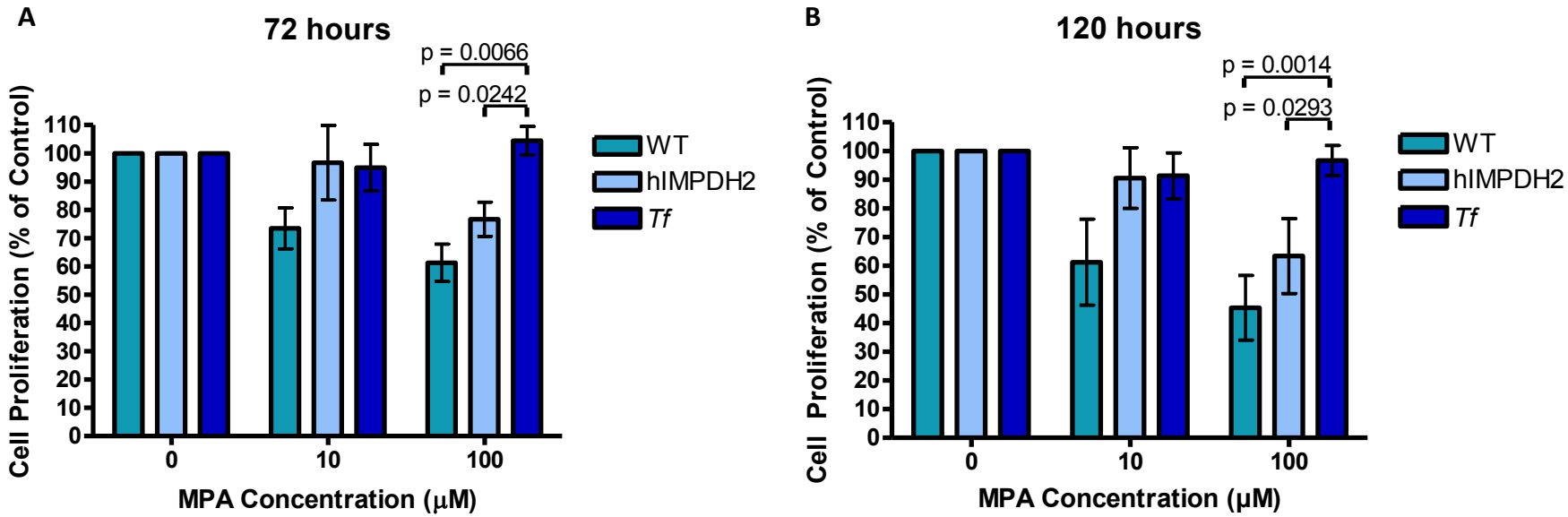


Figure 4.7 Validation of MPA resistance of IMPDH2 variants at 72 and 120 hours. Cells were transfected with either human IMPDH2 (hIMPDH2) or *T. foetus* IMPDH (*Tf*) cDNA expression vectors prior to treatment with 0, 10 or 100 µM of MPA for a period of either A) 72 hours or B) 120 hours. Cell proliferation was determined using a BrdU based ELISA. Results are shown as percentage or untreated control (0 µM MPA). Data represent means of at least 3 independent experiments. Error bars depict standard error of the mean (S.E.M.). A) No significant differences were found between any sample treated with 10 µM MPA. B) No significant differences were found between any sample treated with 10 µM MPA. P-values were calculated using unpaired T-test.

4.3.4 Utilising Salvage Pathway Inhibitors

In order to improve selection of cells carrying MPA-resistant forms of IMPDH, I sought to employ the use of a guanine salvage pathway inhibitor, acyclovir monophosphate. As mentioned above (Section 4.1.2.1), cells have two interconnected pathways to fulfil their guanine nucleotide requirements: *de novo* synthesis and the guanine salvage pathway. IMPDH plays a pivotal role in the *de novo* synthesis pathway, and inhibition of IMPDH via MPA treatment blocks this pathway. Therefore, cells become entirely reliant on the guanine salvage pathway. Inhibition of both pathways should select for MPA resistant cells, provided by the presence of *Tf*-IMPDH.

To investigate the potential of acyclovir monophosphate to modulate the response of *Tf*-IMPDH-expressing cells to MPA treatment, I carried out a BrdU-based cell proliferation ELISA. This time, cells were treated with a combination of 100 μM MPA and a range of acyclovir monophosphate concentrations (0, 1, 10 and 100 μM) for a period of 120 hours, following transfection with *Tf*-IMPDH (or wild type IMPDH2).

Acyclovir monophosphate did not have the predicted effect: increasing the concentration of acyclovir monophosphate (at a constant MPA concentration) did not lead to a decrease in cell proliferation in the wild type cells (Figure 4.8). *Tf*-IMPDH transfected cells demonstrated no significant change in cell proliferation across all concentrations of acyclovir monophosphate tested (ANOVA, $p>0.05$). Overall, there was a modest decrease in the difference between cell proliferation of wild type cells and *Tf*-IMPDH transfected cells, except at 1 μM, where it slightly increases (Table 4.1), though this is not significant (ANOVA, $p>0.05$). In addition, there was no significant difference in cell proliferation across all concentrations of acyclovir monophosphate tested, for both wild type cells and cells transfected with *IMPDH2* cDNA (ANOVA, $p>0.05$). Due to this unexpected response, acyclovir monophosphate co-treatment with MPA failed to improve selection of *Tf*-IMPDH transfected cells.

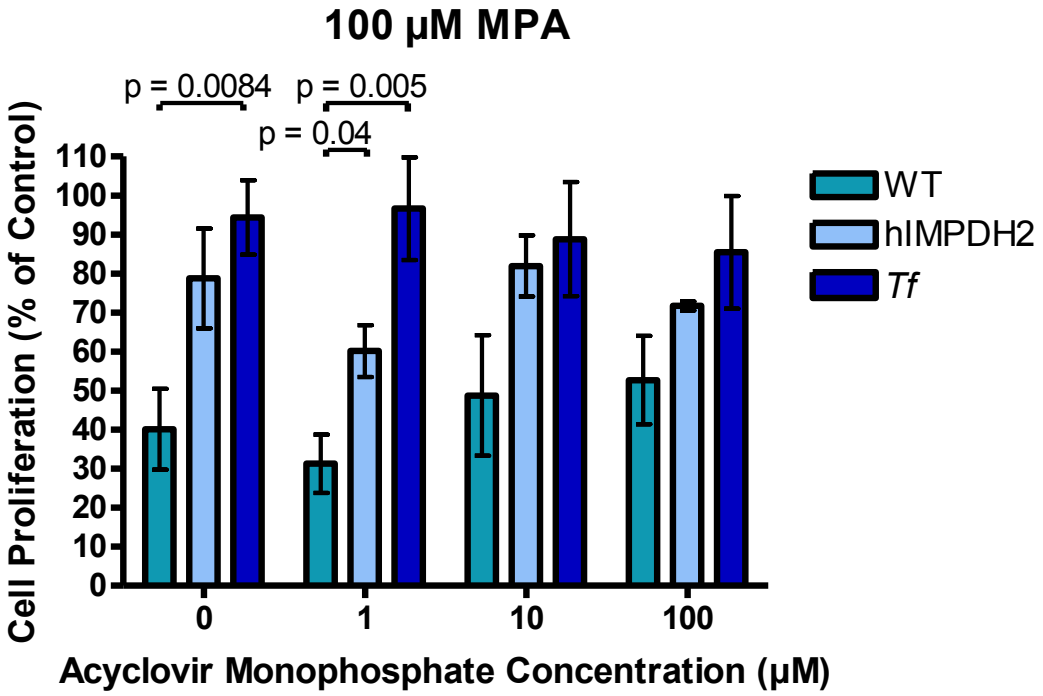


Figure 4.8 Effect of co-treatment with acyclovir monophosphate and MPA. Cells were transfected with either human IMPDH2 (hIMPDH2) or *T. foetus* IMPDH (*Tf*) cDNA expression vectors prior to treatment with varying concentrations of acyclovir monophosphate (0-100 µM) and 100 µM MPA for a period of 120 hours. Cell proliferation was determined using a Brd-U based ELISA. Results are shown as percentage control (DMSO treated). Data represents 3 independent experiments. Error bars depict standard error of the mean (S.E.M.).All significant ($p < 0.05$) differences are indicated (calculated by unpaired T-test).

Table 4.1 The effect of Acyclovir Monophosphate and MPA co-treatment on the Selection of *Tf*-IMPDH. The differences in the levels of cell proliferation between wild type and *Tf*-IMPDH following treatment ± S.E.M. are listed. No statistical significance was found between any of the concentrations, p>0.05 (ANOVA with Bonferroni correction). WT = wild type, *Tf* = *Tf*-IMPDH.

Acyclovir monophosphate (μM)	Absolute difference in cell proliferation between WT and *Tf* (%)
0	54.3 ± 13.8
1	65.3 ± 17.4
10	40.1 ± 17.0
100	32.8 ± 17.6

4.3.5 *IMPDH2* as a Site for Transgene Integration

The development of the IMPDH integration system was not further pursued, in light of the difficulties in successfully selecting MPA-resistant variants from wild type cells. However, it is clear from these experiments that cells can tolerate a degree of inhibition of endogenous IMPDH2. Therefore, it should be possible to use the *IMPDH2* locus as a target for transgene integration, disrupting gene function of one allele, without adverse consequences for cell health.

4.3.5.1 Repair Cassette Design and Construction

I generated an *IMPDH2*-targeting transgene integration cassette by cloning PCR-amplified *IMPDH2* homology arms either side of an *mCherry* transgene under the control of the EF1α promoter. The cassette also contains a puromycin resistance gene downstream of a P2A signal. Figure 4.9 shows the organisation of the repair cassette, which was designed so that any of the six individual sgRNAs could be used to drive its integration. It contains a total of approximately 3.5-Kb homology.

Once constructed and sequence verified, I wanted to ensure that the cassette could express mCherry and that this expression could be enriched by transient puromycin selection. In order to verify this, I carried out transient transfections with the cassette, with and without puromycin selection. I then carried out flow cytometric analysis of mCherry expression (Figure 4.10).

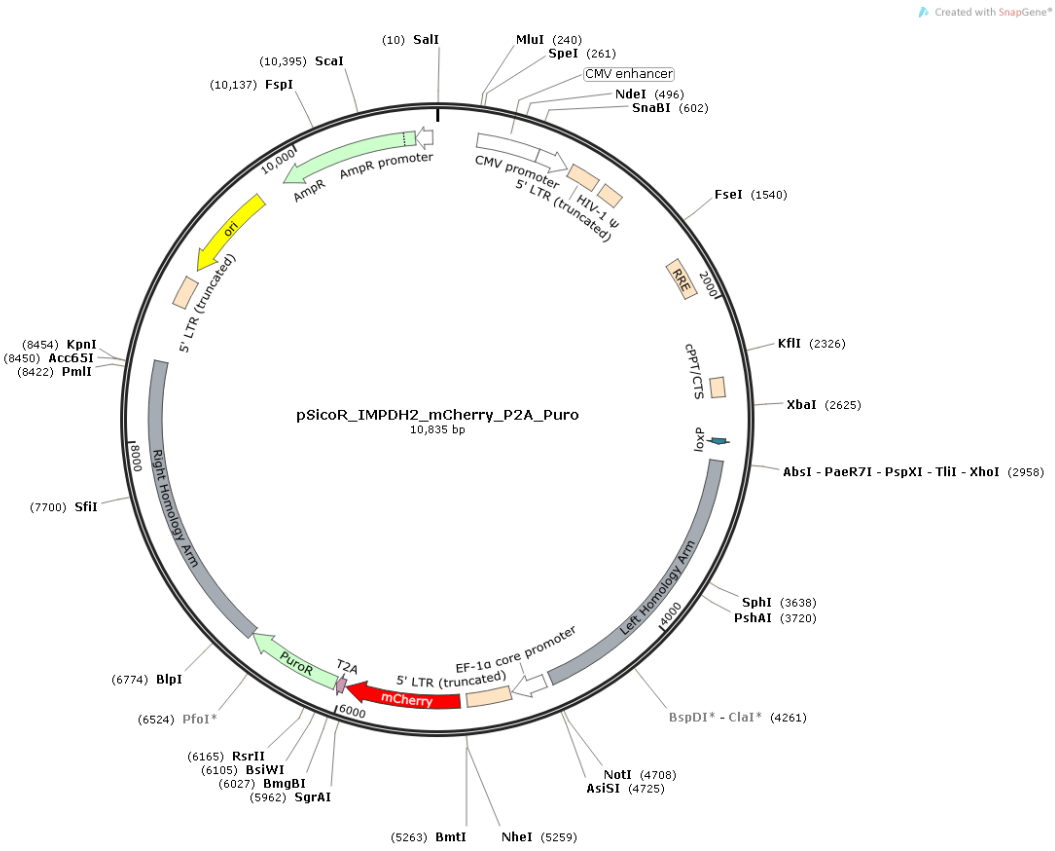


Figure 4.9 *IMPDH2* transgene integration cassette. The cassette has a total homology of ~3.5-Kb, which was designed so that any of the six *IMPDH2*-targeting sgRNAs can be used to drive its integration. The reporter gene *mCherry* is under the control of the EF1α promoter, which also drives the expression of the puromycin resistance gene. The T2A "self-cleaving" peptide sequence allows expression of both genes from the single promoter.

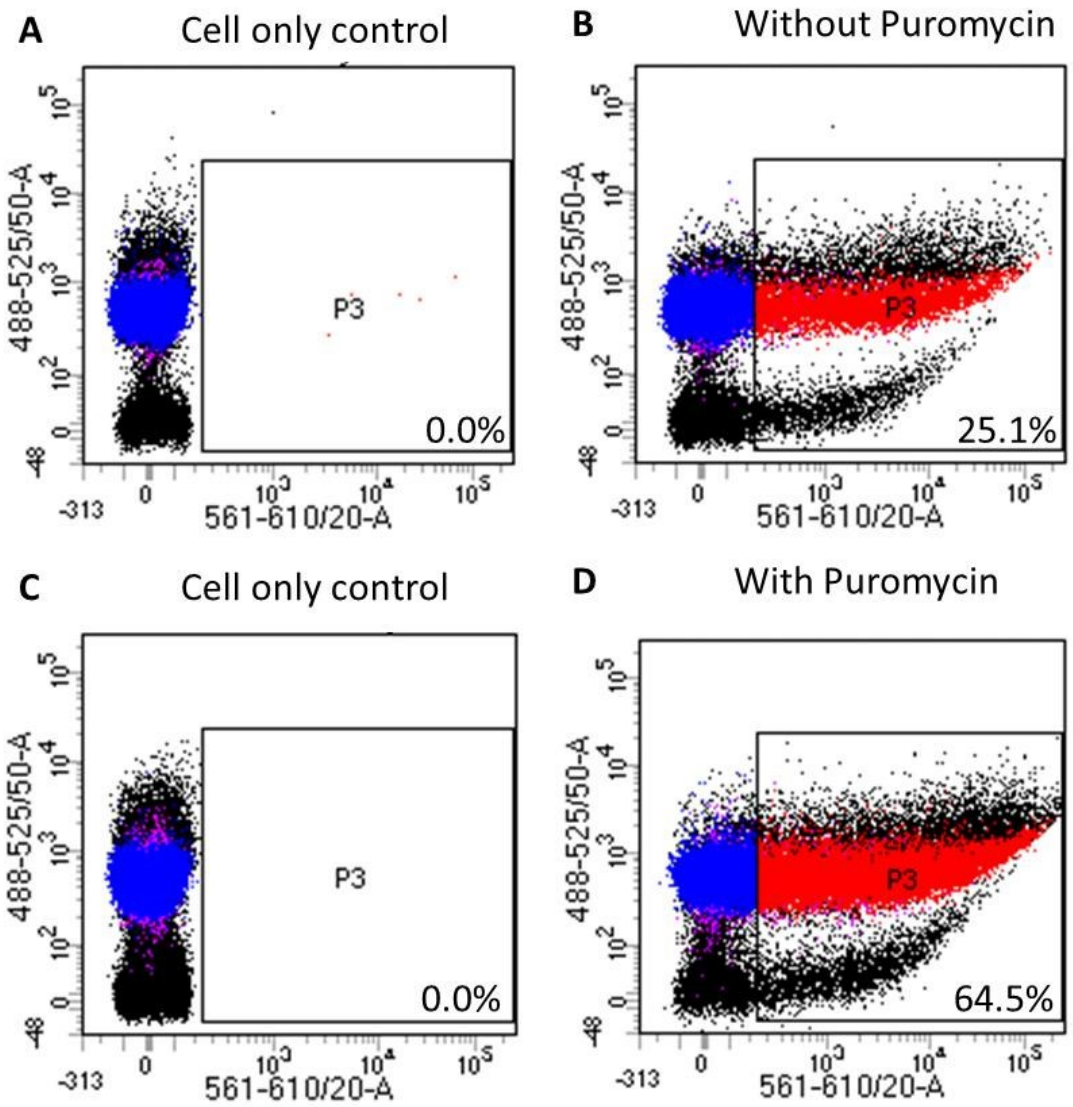


Figure 4.10 Flow cytometric data following transient transfection of cassette, with and without puromycin selection. Scatter plots show mCherry expression levels indicated as percentage of live single cells. Red dots indicate mCherry positive, live single cells; blue dots are live cells; purple dots are live single cells; black dots are dead cells, or cell clumps. For each sample, 50,000 live single cells were analysed. A) Untransfected control for cells not treated with puromycin. B) Cells transfected with repair cassette. C) Untransfected control for cells treated with puromycin. D) Cells transfected with repair cassette and selected with puromycin for a period of 48 hours.

Figure 4.10 indicates that HEK293FT cells transfected with the cassette express mCherry (25.1% of live single cells). In addition, transient puromycin selection increased the proportion of living cells expressing mCherry (64.5% of live single cells).

4.3.5.2 Integration of Repair Cassette

I co-transfected the CRISPR/Cas9 machinery and the repair cassette into HEK293FT cells. I treated these with puromycin for a period of 48 hours to select for transfected cells. Following recovery from puromycin treatment (a further 48 hours), I monitored mCherry expression in each population by flow cytometry (Figure 4.11).

Initial flow cytometric data show that all cells transfected with the repair cassette demonstrate mCherry expression, ranging from 14.5% to 36.3% of living single cells (Figure 4.11). The proportion of mCherry expressing cells rapidly declined over two weeks until less than 1% of live cells were expressing mCherry. At this point I selected for mCherry expressing-populations by FACS.

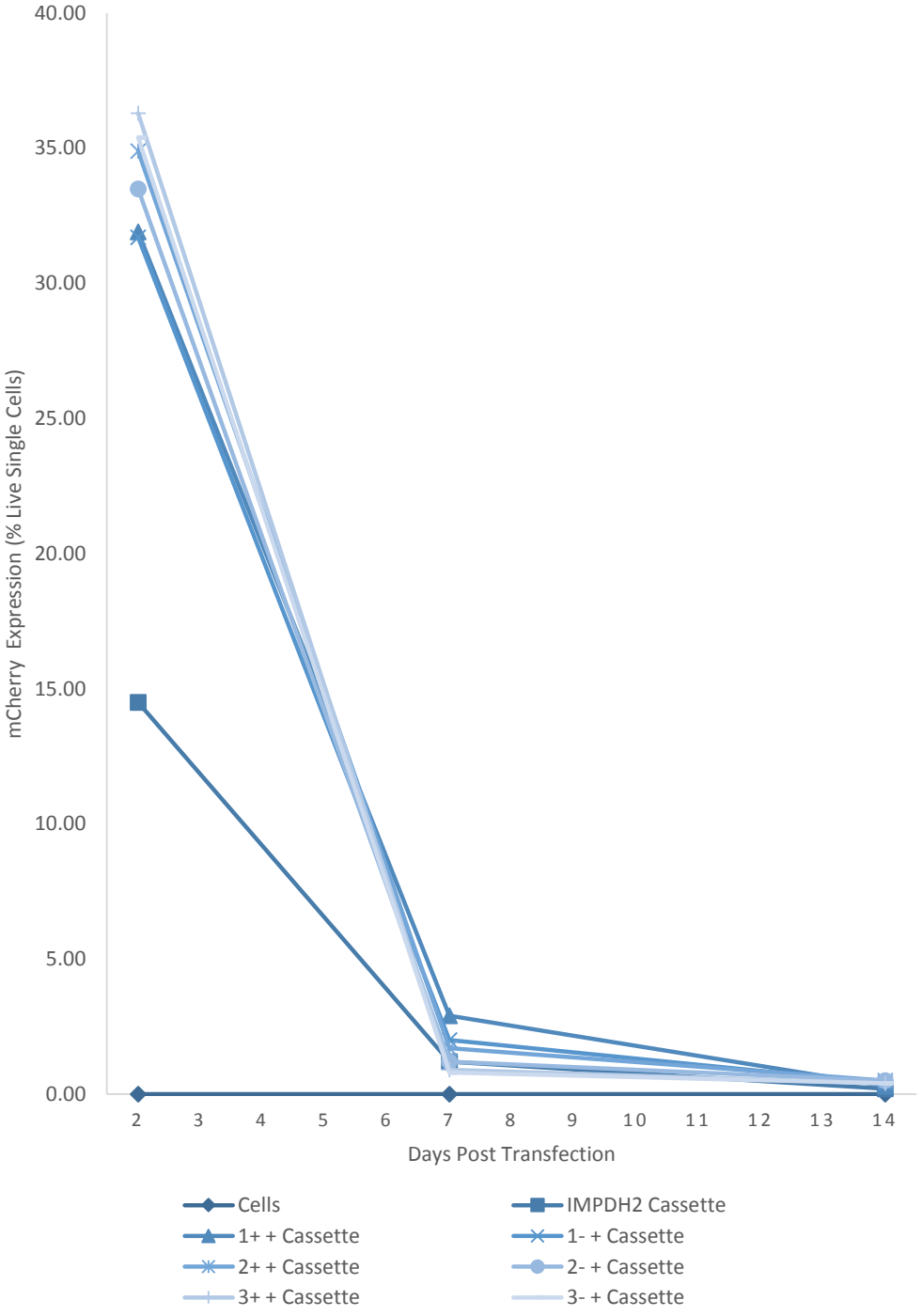


Figure 4.11 Initial mCherry expression in transfected populations. Analysis was carried out on day two following puromycin selection of transfected cells, and then at one week and two weeks post transfection. The expression of mCherry rapidly declined in the first week post transfection. After two weeks, all samples expressed mCherry at <1% live single cells.

4.3.5.3 Generation of Clonal Lines

Following recovery from FACS, I sorted the populations of cells enriched for mCherry expression into 96-well plates (one cell per well) by FACS, based on mCherry expression, in order to isolate clonal lines which retained transgene expression. This resulted in a total of 61 clonal lines, 47 of which were treated with CRISPR/Cas9 and cassette, 6 lines were derived from an mCherry control and the remaining 8 lines were derived from the population transfected with the targeting cassette alone. These were allowed to proliferate until the 6-well plate stage, over a period of 36 days. At this stage, I monitored mCherry expression in each line using flow cytometry (Figure 4.12) and isolated genomic DNA in order to identify clones with correct integration at the *IMPDH2* locus. To identify clonal lines harbouring integration at the *IMPDH2* locus, I carried out a PCR-based assay with locus specific and cassette specific primers; an amplicon of ~2-Kb can only be produced if the cassette has successfully integrated into the *IMPDH2* locus (Figure 4.13).

Figure 4.12 shows the percentage of live cells retaining mCherry expression in each clonal line. These lines are 12 weeks post transfection, thus any observed expression can be considered as long term expression.

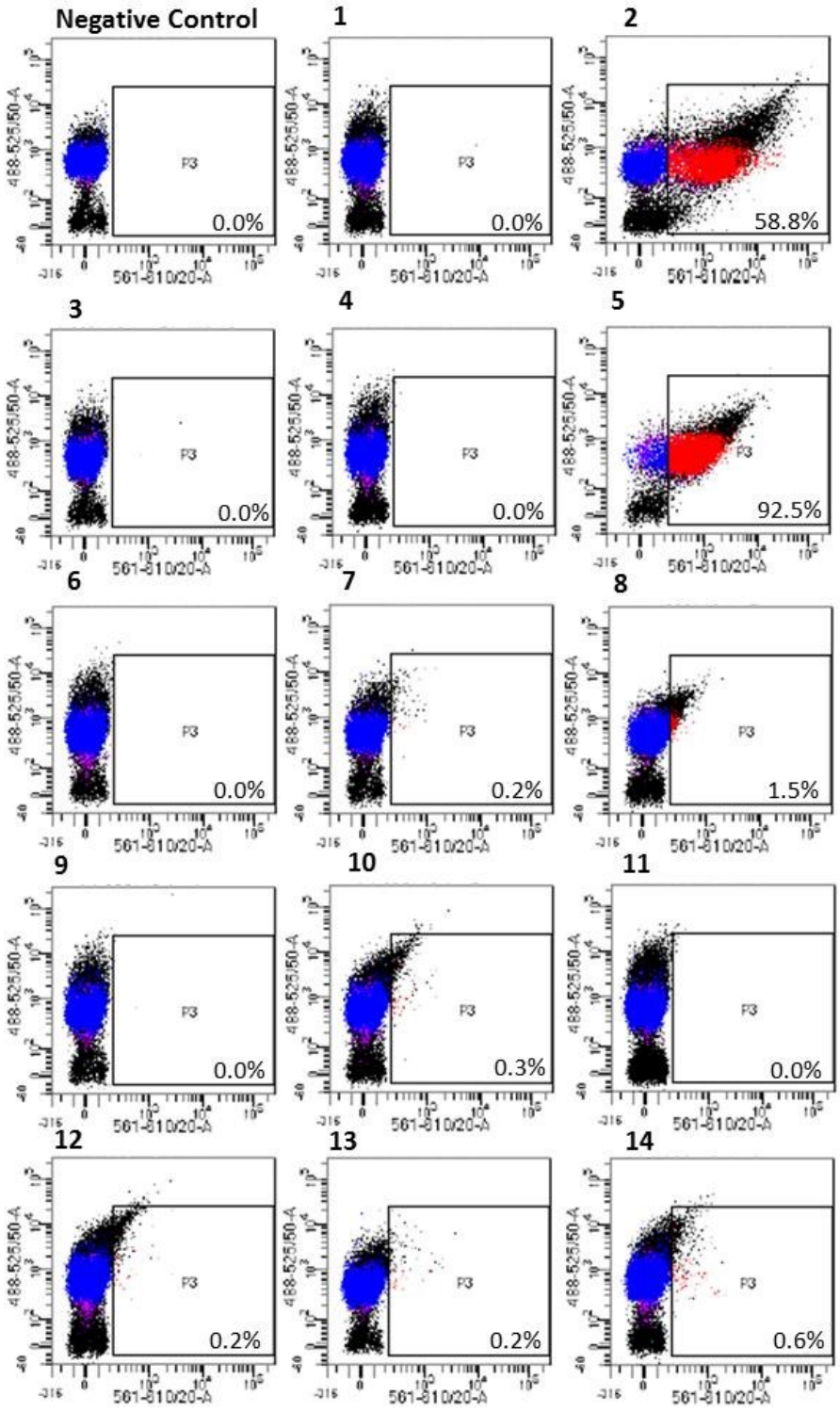


Figure 4.12 Characterisation of clonal lines. Page 1 of 5.

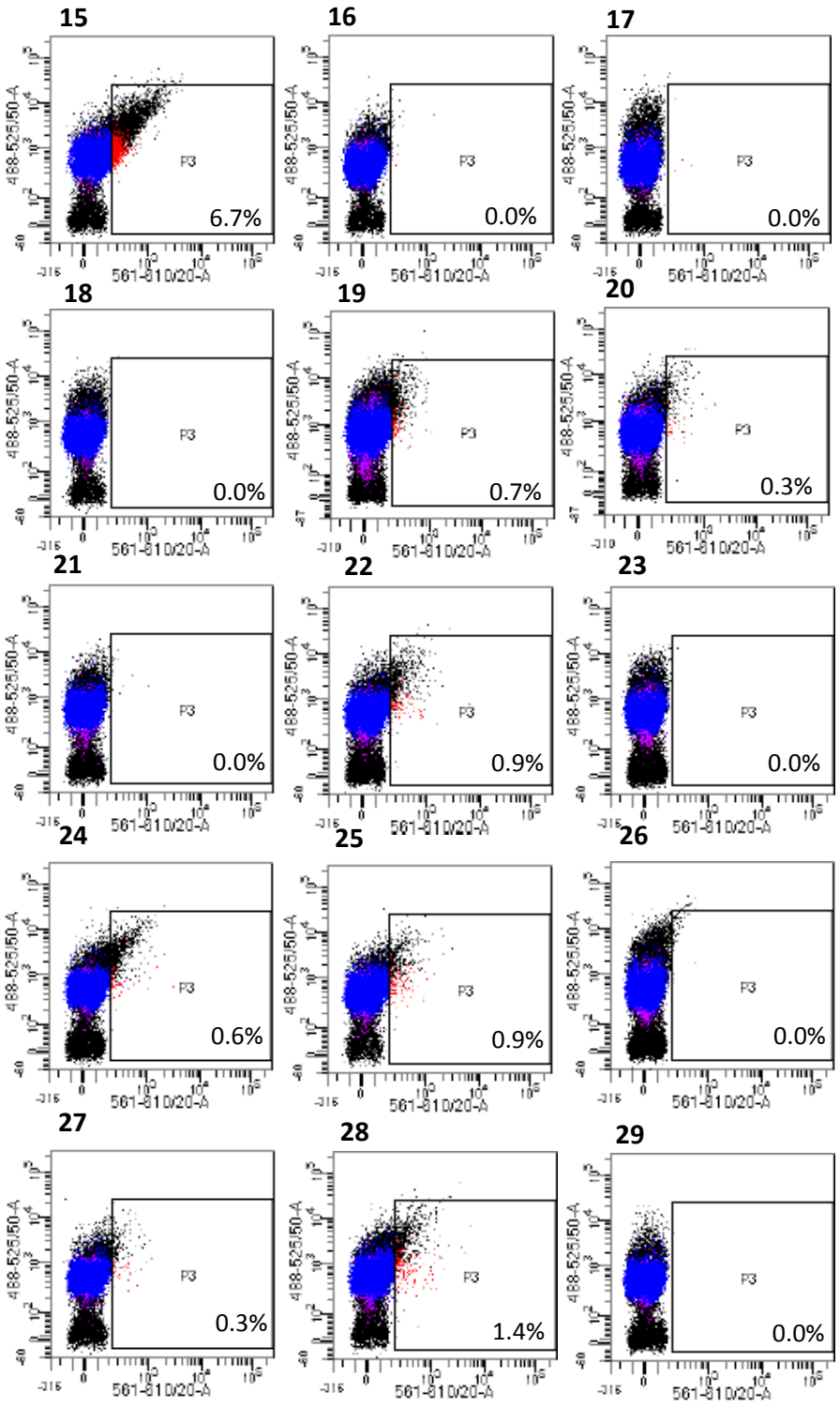


Figure 4.12 Continued. Page 2 of 5.

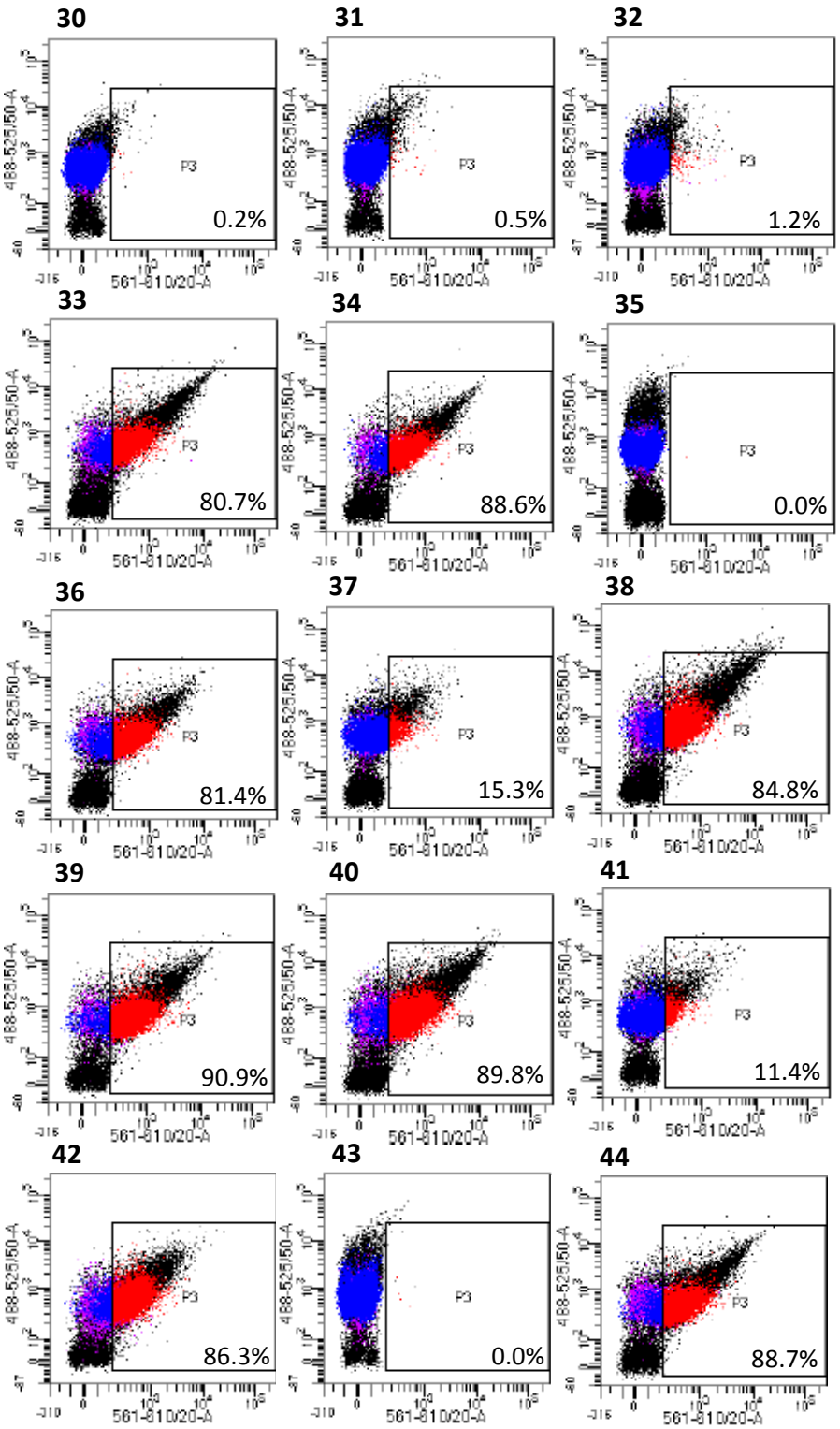


Figure 4.12 Continued. Page 3 of 5.

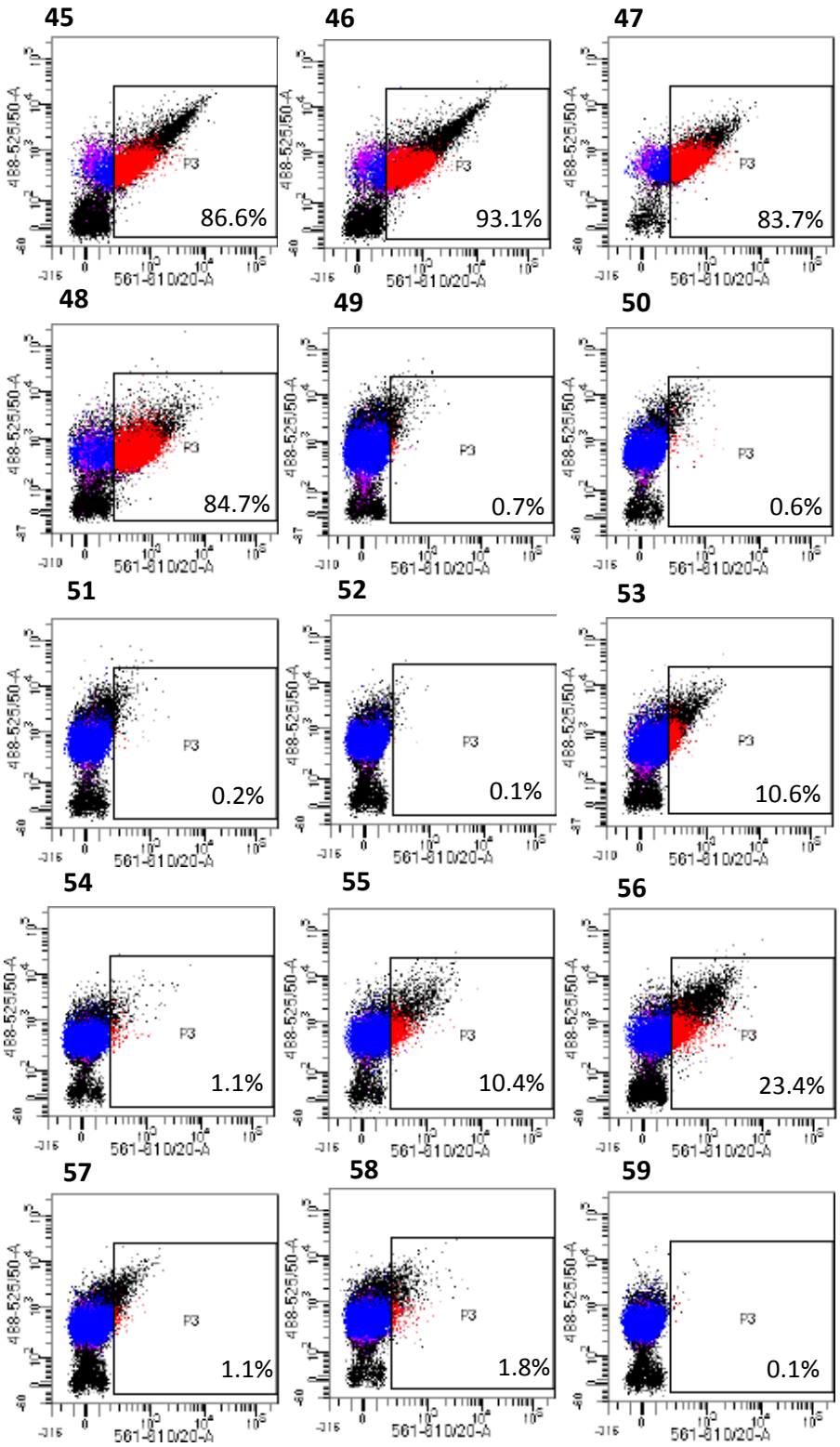


Figure 4.12 Continued. Page 4 of 5.

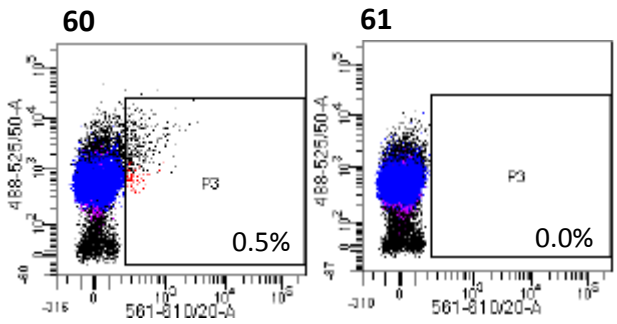


Figure 4.12 Continued. Page 5 of 5. Flow cytometric scatter plots showing mCherry expression levels in each clonal line, indicated as percentage of live single cells. Red dots indicate mCherry positive, live single cells; blue dots are live cells; purple dots are live single cells; black dots are dead cells, or cell clumps. For each clonal line, 50,000 live single cells were analysed. Clones are labelled as their number, 1-61.

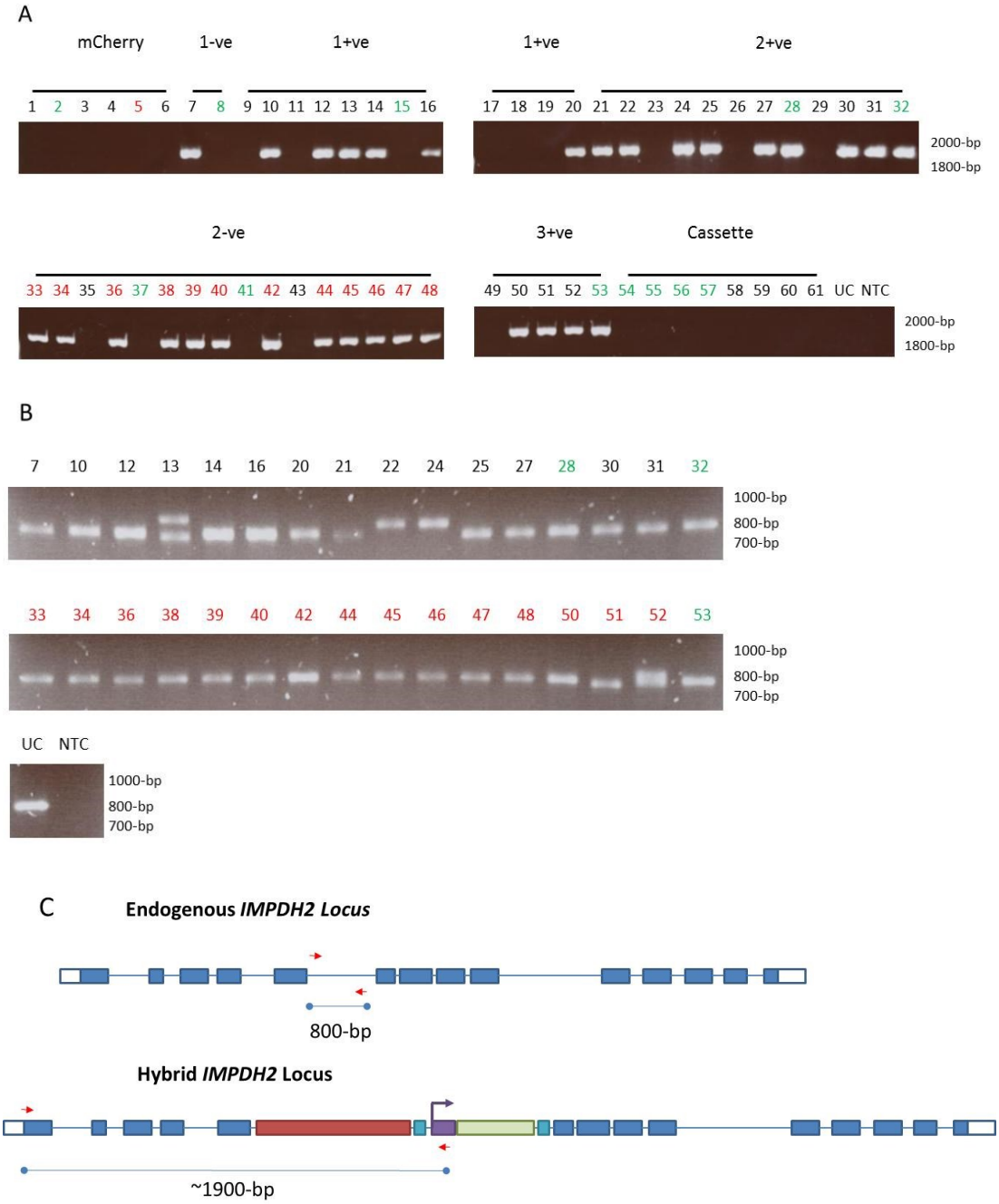


Figure 4.13 A) PCR-based identification of correct integration events at the *IMPDH2* locus. PCR was carried out using locus specific and cassette specific primers; only clones bearing a correctly integrated transgene would produce a product of ~2000-bp. B) PCR-based detection of non-integrated alleles for clones which were positive for cassette integration at *IMPDH2*. Locus specific primers surrounding the target site were used, only alleles without integrated transgene would produce a product of ~1900-bp. C) Schematic of *IMPDH2* locus, indicating location of primers (red arrows) used for integration detection and detection of non-targeted alleles. Predicted PCR product sizes are indicated. Clones are labelled as their number, 1-61. Red numbers indicate clones which had >80% mCherry expression; green numbers indicate mCherry expression of 1-80%. Black numbers indicate <1% mCherry expression. UC = Untransfected control; NTC = No template control.

CRISPR-mediated insertion of the transgene cassette at the *IMPDH2* locus was efficient, with 32 out of 47 (68.1%) CRISPR/Cas9-treated lines demonstrating correct integration (Figure 4.13A). Of these clonal lines harbouring correct integration, 12/32 (37.5%) showed high mCherry expression (>80% of live single cells expressing mCherry) and 3/32 lines showed low expression (1.2%, 1.4% and 10.6% of live single cells expressing mCherry). In addition to these, 11/61 (18%) lines showed mCherry expression without integration at the *IMPDH2* locus, one of these (Clone 5; 1/61; 1.6%) showed high level mCherry expression (95.2% of live single cells expressing mCherry), while 10/61 (16.4%) lines showed low level mCherry expression (mean mCherry expression of 13.2% ± 17.6). All lines with integration at *IMPDH2* showed monoallelic integration, as demonstrated by the presence of an approximately 1900-bp band following PCR with locus specific primers (Figure 4.13B). In some lines, particularly clone 13 (derived from sgRNA_1+), there is evidence of indels on the non-targeted allele (indicated by the presence of additional bands). Clones 22, 24 and 51 also showed evidence of indels, indicated by the presence of a single band, larger than the expected band (Figure 4.13B).

4.4 Discussion

While in principle promising, the drug-selectable transgene integration system proposed here proved to be unfeasible in practice. This was mainly due to the high levels of residual survival observed in HEK293FT cells, following treatment with MPA. However, through initial experiments, I have for the first time, been able to demonstrate the potential of *IMPDH2* as a novel site for efficient transgene integration and its ability to support robust long term expression.

4.4.1 Targeting *IMPDH2* with CRISPR/Cas9

As demonstrated in Chapter 3, the sgRNAs I designed to target *IMPDH2* are all active with wild type Cas9 (as determined by the SURVEYOR assay). I confirmed this by sequencing individual amplicons from clonal lines. The range of indels observed is consistent with that demonstrated by other groups (Shen et al. 2014; Hruscha et al. 2013; Cong et al. 2013; Mali, Yang, et al. 2013) and confirms that

these tend to cluster around the predicted cleavage site – 3-bp upstream of the PAM. No indels were observed for any of the sgRNAs when provided as a pair along with the Cas9n, consistent with the results observed with the SURVEYOR assay (Chapter 3). This corroborates reports suggesting that a tail-to-tail orientation of the paired sgRNAs is favourable (Shen et al. 2014), as these pairs were all orientated in a head-to-head configuration.

In some cases more than two mutant alleles were observed in each clonal line. This may be due to the presence of at least three *IMPDH2* alleles in HEK293FT cells, which is unsurprising as these cells are hypotriploid. However, these could also result from errors during cloning; FACS is not infallible so some lines could have inadvertently resulted from two cells rather than the intended single cell, despite careful screening for this during clonal line production.

4.4.2 Variable Response Following MPA Treatment

One of the major hurdles in creating the drug selectable transgene integration system was the variable response observed following treatment of cells with MPA, particularly the high residual survival of wild type cells, ranging from ~40% to ~60% of control. This was higher than anticipated based on the results of Qasim and colleagues (2011), who showed that HEK293 cells (the parent line of HEK293FT cells) had ~10% survival following treatment with MPA. Indeed, I was unable to replicate this under the same experimental conditions, observing a mean of ~40% survival, following three independent experiments. It is unclear why I consistently observed residual proliferation here. A potential explanation is batch variability in drug concentration though it seems unlikely that such variation would have been large, and would not account for the intra-experiment variability that I also observed. It is also a possibility that inter-batch variability in the antibody used in the ELISA caused some of the observed differences in residual survival. In addition, potential differences in seeding density between experiments may have resulted in apparent differences in survival.

However, the response following treatment of transfected cells was more consistent. I would expect that, if variations in drug concentration, ELISA reagents, or cell

seeding density had occurred, then there would be more variation in transfected cell response to MPA, mirroring that observed of untransfected (wild type) cells.

The overall high level of residual survival following MPA treatment indicates that the salvage pathway is likely to play a significant role in the provision of guanine nucleotides in HEK293FT cells. Treatment with MPA leads to inhibition of the *de novo* synthesis pathway so cells become reliant on the salvage pathway to provide their guanine nucleotide needs. However, treatment with MPA may select for cells in which the salvage pathway is more active, and this may explain some of the variation seen in wild type response to MPA. This may be masked in transfected cells which can bypass inhibition via the presence of the IMPDH variant. This, in part, may explain the difference in variability of response between transfected and wild type cells.

4.4.3 IMPDH^{IY} did not Confer Increased MPA Resistance

Interestingly, my findings for IMPDH^{IY} do not replicate those of others in the field. Initial studies in MPA-resistant neuroblastoma lines, in which the variant was identified, also indicated that gene amplification was present, which could account for some, though not all, of the reported increase in MPA resistance (Hodges et al. 1989). However, in their study, Sintchak et al. (1996) showed that the Thr333Ile mutation alone can increase MPA-resistance 300-fold compared to wild type. It would therefore be worth investigating whether the Ser351Tyr mutation may in some way mitigate the resistance conferred by the Thr333Ile substitution alone. Indeed, the activity of IMPDH^{IY} is less than the wild type IMPDH (Farazi et al. 1997) and so this may account for its inability to provide increased resistance in this system and may explain why it performed worse than overexpression of IMPDH2 here (Figure 4.6). However, it has also been shown by others that overexpression of the IMPDH^{IY} variant can confer resistance to MPA above wild type (Sangiolo et al. 2007; Schroeder et al. 2006). This disparity may in part be due to cell line differences and perhaps the relative activity of the salvage pathway, particularly as these studies were conducted in T lymphocytes, which have little or no salvage pathway activity (Allison et al. 1977).

4.4.4 Salvage Pathway Inhibition

Early biochemical studies indicated the potential of acyclovir monophosphate to inhibit HPRT (Tuttle et al. 1983), the key enzyme involved in the guanine salvage pathway (Figure 4.2). It was also demonstrated to inhibit purine nucleoside phosphorylase (PNP), another enzyme involved in the salvage pathway (Tuttle & Krenitsky 1984). However, co-treatment with acyclovir monophosphate and MPA did not improve selection of *Tf*-IMPDH. In fact, acyclovir monophosphate did not significantly affect cell proliferation in any treatment group. The reasons for this are not clear and further studies, which are beyond the scope of this thesis, would be required to identify the underlying mechanisms.

4.4.5 *IMPDH2* Supports Transgene Integration and Expression

In order to demonstrate *IMPDH2* as a suitable site for conventional site-specific transgene integration, I designed a targeting cassette which could be incorporated following cleavage by CRISPR/Cas9. For initial proof of principle studies, I used an *mCherry* gene under the control of the EF1α promoter so that I could easily monitor expression of the transgene once integrated at *IMPDH2*.

The data presented here demonstrate the ability of *IMPDH2* to support robust long term expression of transgenes following CRISPR/Cas9-mediated integration of the transgene cassette; 37.5% of integrated clones showed greater than 80% mCherry expression after 3 months. As far as I am aware this is the first time *IMPDH2* has been demonstrated as a site for transgene integration. These data indicate that the CRISPR/Cas9 system can be used to efficiently target integration of transgenes (from plasmid DNA). The efficiency achieved in this study (68.1% integration) is comparable with those demonstrated by other groups in cultured human cells (Byrne et al. 2014) and other model systems (Böttcher et al. 2014; Gratz et al. 2014; Yang et al. 2013). However, as this work was carried out in cultured cells, it is not necessarily predictive for expression in animals or humans.

Integration efficiency varied depending on the sgRNA used to drive Cas9 cleavage. The total number of clones generated from each individual sgRNA also showed variability, so it is difficult to make any definitive conclusions. However, the fact

that different numbers of clones were generated may itself indicate the usefulness of each sgRNA for targeting transgene integration. Indeed, all clones with correct integration demonstrating robust mCherry expression were generated with sgRNA_2-. The reason why this particular sgRNA should be more efficient at driving transgene integration is not clear, though it is common for multiple sgRNAs targeting the same locus to show different activities

In all cases, integration at *IMPDH2* appeared to be monoallelic (Figure 4.13B). Due to the complex karyotype of HEK293FT cells, it is possible that integration occurred at more than one allele, but it remains true that at least one allele with no integration is present in all targeted clones. It appears that whilst integration is not biallelic, there is evidence that targeting of the non-integrated alleles has occurred. In one case (clone 13) it appears that an insertion event has occurred in a non-targeted allele, while clones 22 and 24 only show presence of a larger PCR amplicon (suggestive of an insertion event at the non-targeted allele). This is similar to the 237-bp insertion I observed previously in clonal lines (Section 4.2.1.2).

Many established cell lines, commonly used in the laboratory, have abnormal karyotypes often possessing more or less than the diploid number of chromosomes (Table 4.2). This could potentially complicate the production of transgenic cell lines. However, in cases where high level protein expression is desirable (for example, monoclonal antibody production), the ability to integrate two copies whilst preserving a wild type allele may be beneficial. However, in cell lines where the chromosomal number is less than diploid, it may be necessary to establish the copy number of the target locus, as disruption of a single copy locus may be deleterious, particularly in the case of *IMPDH2,* if activity of the salvage pathway is essential for cell survival.

Table 4.4.2 Ploidy of commonly used immortalised cell lines. Data collated from ATCC, Sigma-Aldrich® and Creative Bioarray.

			Chromosome number	
Cell Line	**Organism**	**Tissue of Origin**	**Modal**	**Diploid**
HEK 293	*Homo sapiens*	Embryonic kidney	64	46
HeLa	*Homo sapiens*	Cervical epithelium	82	46
A549	*Homo sapiens*	Lung epithelium	66	46
RPE1	*Homo sapiens*	Retinal pigmented epithelium	46	46
COS-7	*Cercopithecus aethiops*	Kidney	57	60
SH-SY5Y	*Homo sapiens*	Bone marrow	47	46
Vero	*Cercopithecus aethiops*	Kidney epithelium	58	60
HCT116	*Homo sapiens*	Colorectal carcinoma	45	46
MCF7	*Homo sapiens*	Mammary gland	82	46
CHO-K1	*Cricetulus griseus*	Ovary	20	22
HEP-G2	*Homo sapiens*	Liver	52	46
CACO-2	*Homo sapiens*	Colon	96	46

Rapid decline in the proportion of mCherry-expressing cells occurred following transfection (Figure 4.11). This is likely a result of the loss of the plasmid DNA, while later loss of expression may be attributed to silencing of integrated copies, or outgrowth of cells harbouring integrated transgenes by cells without transgene integration.

Some of the correctly integrated clones demonstrated a variegated phenotype, which was also observed in mCherry-expressing lines without correct integration. Indeed, clones 28, 32 and 53 showed correct integration but exhibited low levels of mCherry expression (1.4%, 1.2% and 10.6% respectively). In addition, 16 lines show correct integration but did not express mCherry (Figure 4.12B). Variegated expression of transgenes is predominantly caused by position effects. These can result from integration of transgenes in or close to heterochromatic regions, such as those found near telomeres and centromeres. Under these circumstances, heterochromatin can spread to the transgenes and lead to extinction of their expression (Decottignies 2014). Transfected plasmid DNA becomes linearized upon entry to the nucleus and can be ligated to form concatemeric arrays which become integrated into the genome (Wurm 2004); their expression is often silenced by repeat-induced gene silencing

mechanisms (Garrick et al. 1998). It is also possible that mutations or rearrangements have occurred in the transgenes during integration which have rendered them inactive and may explain why some lines show no expression despite correct integration at *IMPDH2*, though further studies would be required to confirm this. The presence of variegation in some clones implies that integration at *IMPDH2* does not completely prevent partial silencing of the transgene, though the results obtained here suggest that this is less common upon integration at *IMPDH2* than when the transgene is integrated randomly. Further studies are also required to identify the underlying mechanisms of the variegated phenotype observed in some clones, and to fully characterise *IMPDH2* as a site for transgene integration and expression.

4.5 Future Perspectives

The results presented here demonstrate the potential of *IMPDH2* to enable efficient transgene integration –mediated by CRISPR/Cas9 and support robust long term expression of transgenes. While showing initial promise, it will be necessary to further investigate the potential of *IMPDH2* for this purpose. Firstly, I think it prudent to generate further clonal lines in order to validate the results obtained here, and investigate if the integration efficiencies achieved by each sgRNA can be replicated. Should this be the case, I think it wise to perform future experiments with sgRNA_2-ve which showed the greatest promise here, both in terms of integration efficiency and in robustness of subsequent transgene expression.

In addition, I think it would be important to further characterise the clonal lines generated in this study. Firstly, it would be useful to demonstrate that correct integration is single copy and that other random integration events have not concurrently occurred. It would also be necessary to sequence integrated transgenes to identify any rearrangements or mutations which could have occurred during the integration process. This would be particularly important for lines which show variegated or no expression. It would also be worth identifying any markers of transgene silencing, which may be present, for example DNA methylation or histone modifications at the site of the transgene integration. For clones which do not demonstrate correct integration, the site of transgene integration should be identified,

which may prove to be useful in future experiments as sites for transgene integration and expression, especially in clones such as clone 5 that showed high level mCherry expression.

Next, I think that the protocol could be further optimised. It should be possible to generate clonal lines immediately following transfection. This could be carried out in the presence of puromycin selection to investigate whether this would have any effect on the proportions of correctly integrated and mCherry-expressing clones.

A recent study has demonstrated that small molecules , specifically Brefeldin A and L755507, could increase the integration efficiency for large transgenes by up to threefold (Yu et al. 2015). It would be useful to investigate if these molecules could further increase the efficiency of targeting at the *IMPDH2* locus.

Once further optimised, it would be prudent to use a transgene of therapeutic or biological interest, rather than a reporter gene, to investigate any gene specific factors which affect either the efficiency of integration or the robustness of long term expression, and demonstrate the worth of this locus for wider applications.

Chapter 5 Development of an *EEF1A1* Knockout Cell Line

5.1 Introduction

The genome editing revolution brought about by the availability of SSNs, particularly TALENs and the CRISPR/Cas9 system, has great potential for improving the way in which we study gene function. Specifically, it has greatly improved the ease of creating gene knockouts in many model systems (Chu et al. 2015; Wettstein et al. 2015; Zhou et al. 2015; Han et al. 2014; Shalem et al. 2014b; Yang et al. 2014). One case where this is useful is when studying almost identical proteins; for example, the two isoforms of the eukaryotic translation elongation factor 1 alpha (eEF1A), eEF1A1 and eEF1A2, which are 92% identical and 98% similar at the amino acid level. The eEF1A1 isoform is expressed almost ubiquitously except in a few tissues in which it is replaced by the eEF1A2 isoform during early postnatal development. It remains unclear why this switch occurs and whether other functions, in addition to their role in protein synthesis, unique to each isoform may be responsible. In order to study the roles of each isoform in more detail, I aimed to create a knockout cell line, which would only express the eEF1A2 isoform, using CRISPR/Cas9 technology.

5.1.1 Role in Eukaryotic Translation

The translation of mRNA into protein is an essential cellular process; as such, mediators of this process are key players in cellular survival. Translation is organised into three phases, initiation, elongation and termination, each requiring the action of several proteins, known as translation factors.

During the elongation phase, aa-tRNAs are delivered to the ribosome to provide amino acids to the growing polypeptide chain (Figure 5.1). This process is facilitated by the action of eukaryotic elongation factors (eEFs); eEF1A is a key mediator of this process responsible for delivering aa-tRNAs to the ribosome (Figure 5.1).

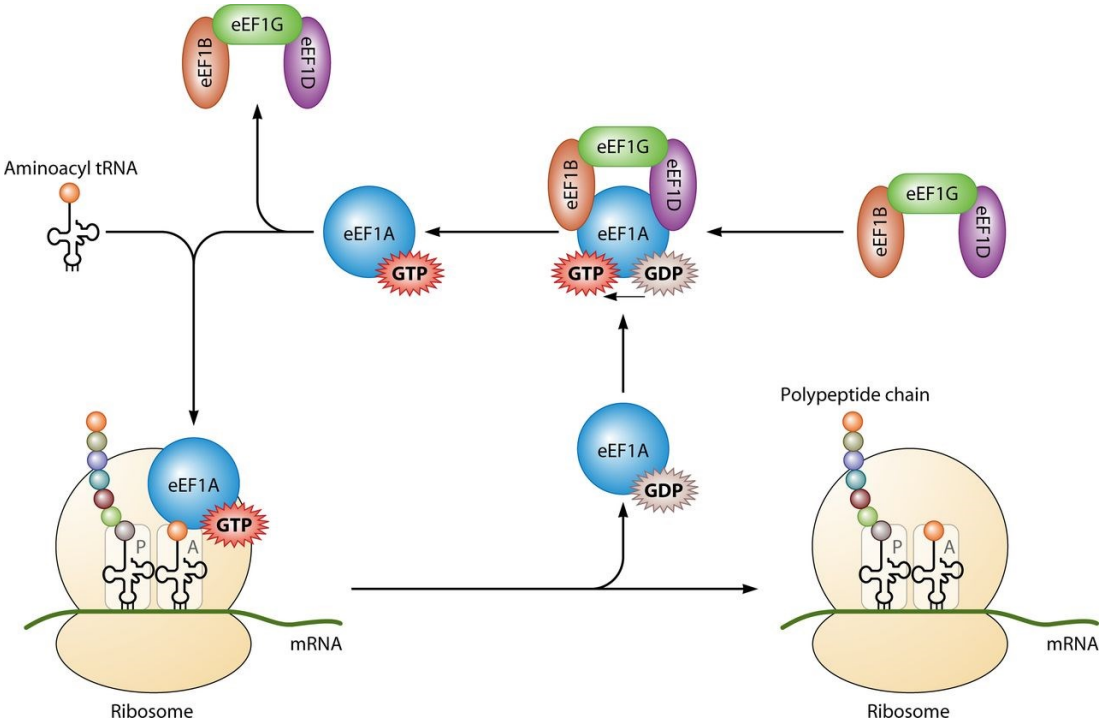


Figure 5.1 Translation Elongation (not to scale). Translation elongation is mediated by eukaryotic translation elongation factors (eEFs). During elongation, aa-tRNAs are delivered to the ribosome, so that amino acids can be added to the growing polypeptide chain. Firstly, eEF1A binds GTP and then associates with an aa-tRNA molecule; eEF1A is responsible for delivery of the aa-tRNA to the ribosomal A site. Once the aa-tRNA is bound, eEF1A hydrolyses GTP and the free eEF1A-GDP complex is released from the ribosome. The eEF1A-GDP complex then associates with the eEF1B complex (composed of eEF1B, eEF1D and eEF1G) which is responsible for the exchange of GDP for GTP, allowing the eEF1A-GTP to associate with another aa-tRNA and the cycle to begin again. Once bound to the A site, the aa-tRNA is translocated to the P site by the action of eEF2, requiring the energy from GTP hydrolysis. Once here the amino acid forms a peptide bond with the last amino acid on the growing polypeptide chain. During translocation, the tRNA at the P site (which has delivered its amino acid) is translocated to the E site where it is released and can be recycled. Each tRNA molecule has an anticodon, which is complementary to the codon in the mRNA molecule and specifies the appropriate amino acid; only complementary tRNAs (carrying the respective amino acid) will be able to bind, thus ensuring the fidelity of the growing polypeptide chain. Taken with permission from Li et al. (2013).

5.1.2 eEF1A Isoforms

5.1.2.1 *EEF1A1* and *EEF1A2* Genes

While there are several *EEF1A* genes in the mammalian genome, only *EEF1A1* and *EEF1A2* are known to be protein coding, encoding the two mammalian isoforms, eEF1A1 and eEF1A2, respectively. The other genes are likely to be retropseudogenes originating from *EEF1A1* (Lund et al. 1996; Lee et al. 1993; Madsen et al. 1990).

The *EEF1A1* gene has been mapped to chromosome 6q14 while *EEF1A2* is located on chromosome 20q13 (Lund et al. 1996). Both genes have eight exons and seven introns, with highly similar coding regions, 75% identical at the nucleotide level (Knudsen et al. 1993), and conserved intron exon boundaries (Bischoff et al. 2000). However, the introns, promoters and untranslated regions (UTRs) show little similarity (Bischoff et al. 2000; Knudsen et al. 1993). The *EEF1A1* gene spans approximately 3.5-Kb, with a strong promoter, TATA box and putative Sp1 and AP1 transcription factor binding sites (Uetsuki et al. 1989). On the other hand, *EEF1A2* spans ~12-Kb, including a 2-Kb promoter region (Bischoff et al. 2000); this disparity in size is largely due to the presence of larger introns in *EEF1A2*. *EEF1A2* lacks a TATA box and a 5' terminal oligopyrimidine tract (following the CAP structure) present in *EEF1A1* (Bischoff et al. 2000). Furthermore, *EEF1A2* has an enhancer box (E-box), EGR (early growth response) transcription factor binding site and a myocyte enhancer factor 2 (MEF2) site, not present in *EEF1A1* (Bischoff et al. 2000), suggestive of their independent regulation.

5.1.2.2 eEF1A1 and eEF1A2 Proteins

The two isoforms are 92% identical and 98% similar at the amino acid level (Figure 5.2), and have highly similar structures (Soares et al. 2009).

The two isoforms show near identical translation efficiencies (Kahns 1998) and the residues involved in eEF1B and aa-tRNA binding are completely conserved between both isoforms (Soares et al. 2009). However, they differ in their relative affinities for GDP and GTP; eEF1A1 binds GTP more strongly than GDP, while for eEF1A2 the opposite is true (Kahns 1998), though the residues in contact with GTP/GDP are completely conserved between the two isoforms (Soares et al. 2009). However,

residues in close proximity to the guanine-binding pocket are isoform specific, in particular Asn197/His197; in eEF1A1, Asn197 can form an H-bond with Asp156, but such an interaction cannot occur in eEF1A2 between Asp156 and His197 which may account for differences observed in GTP/GDP binding (Soares et al. 2009).

In addition, many of the residues which differ between the two isoforms can be differentially post-translationally modified (Soares & Abbott 2013); acetylation, phosphorylation, S-nitrosylation and ubiquitination can occur in an isoform specific manner (Soares & Abbott 2013). This has led to the hypothesis that differences in the post-translational modifications of the isoforms can accentuate the structural differences. Indeed, some of the sites that can be post-translationally modified lie within known functional regions, including the eEF1B-binding site, which may influence the function of each variant. In addition, many of the differentially modified residues lie within the two clusters of sequence variation (Soares & Abbott 2013), potentially indicating a diverged role for each isoform.

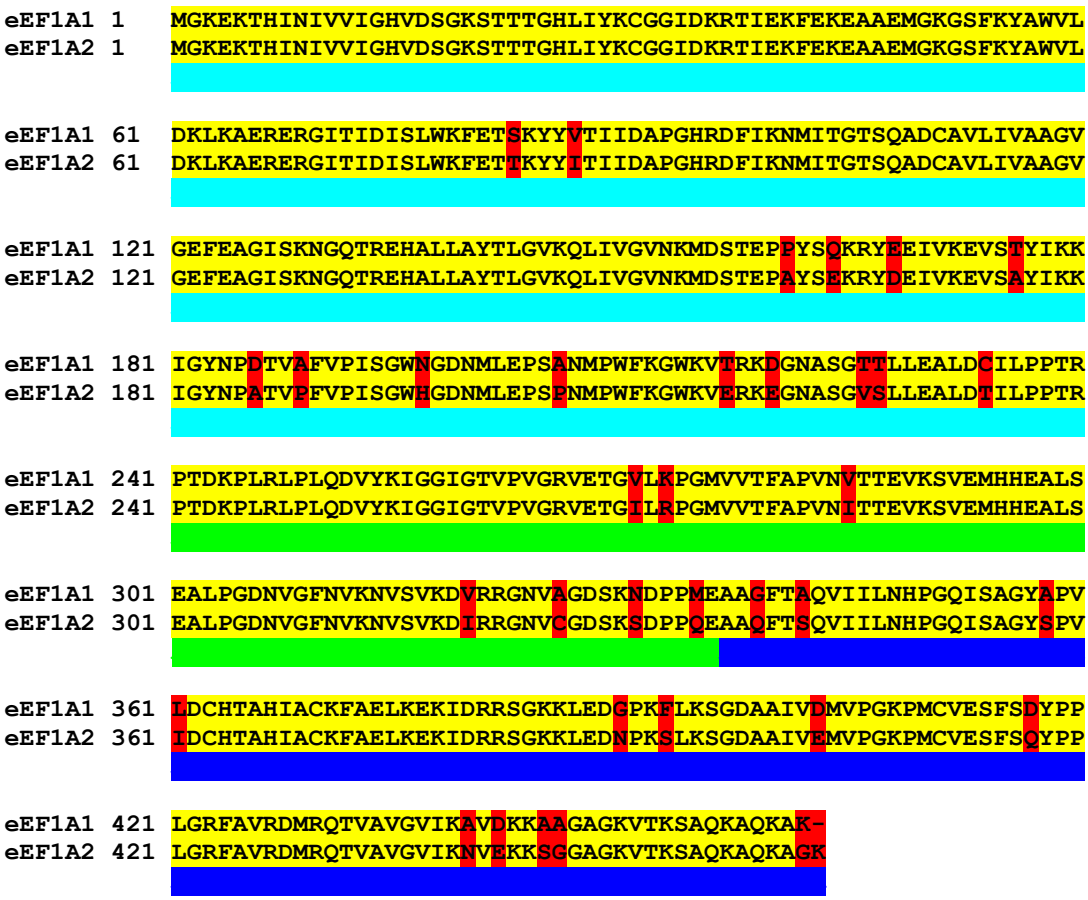


Figure 5.2 Pairwise alignment of eEF1A1 and eEF1A2 amino acid sequences. The protein is organised into three domains, domains I, II, and III. There are a total of 36 amino acid differences between the two isoforms and 32 of these are found throughout the three domains; the remaining four are found within the C-terminal disordered region (Soares et al. 2009). The majority of these variable residues form two distinct clusters on one face of the protein – the opposite face to the C-terminal eEF1B-binding site (Soares et al. 2009). Matching amino acids are highlighted yellow. Residues that differ are highlighted red. Domain boundaries are indicated (Domain I: Cyan; Domain II: Green; Domain III: Dark blue). Adapted from Soares et al. (2009).

5.1.2.3 Two Isoforms to Perform One Role

Even though the two isoforms are highly similar in their structures (Soares et al. 2009) and translation abilities (Kahns 1998), they show non-overlapping expression patterns. The eEF1A1 isoform is almost ubiquitously expressed, while expression of eEF1A2 is limited to adult skeletal muscle, heart, large motor neurones, pancreatic islet cells and enteroendocrine cells in the gut (Newbery et al. 2007; Knudsen et al. 1993). This pattern of expression is laid down during development; eEF1A1 is expressed ubiquitously in the foetus but is replaced by eEF1A2 in the aforementioned tissues during postnatal development. In the mouse this switch has completed by postnatal day 21 (Newbery et al. 2007). The reasons for this switch remain unclear, though may be due to differences in the activity profiles of each isoform.

5.1.2.3.1 Moonlighting Functions

In addition to their role in protein synthesis, the two isoforms have been implicated in wide ranging cellular processes, which are often unique to each isoform.

The eEF1A1 isoform is known to bind and bundle actin and microtubules (Murray et al. 1996; Shiina et al. 1994). Though this has not been confirmed for the eEF1A2 isoform, which may hint at a reason why it is expressed – rather than eEF1A1 – in certain cells. These cells tend to have a very stable architecture, such as neurones, so expression of eEF1A1 may be deleterious.

It has been speculated that eEF1A also plays a role in proteolysis as it is involved in protein quality control via translation-coupled protein degradation (Mateyak & Kinzy 2010). Studies *in vitro* have demonstrated its involvement in degradation of N alpha-acetylated proteins (Gonen et al. 1994) and its ability to facilitate folding of denatured proteins (Hotokezaka et al. 2002). UV cross-linking studies have demonstrated the interaction of eEF1A with misfolded proteins, but not correctly folded proteins, and revealed its interaction with the polypeptide following release from the ribosome (Hotokezaka et al. 2002).

It has emerged that eEF1A1 is involved in eliciting the heat-shock response via association with heat shock RNA 1 (HSR1) and together are essential for the

trimerisation of monomeric heat shock factor 1 (HSF1). This allows movement to the nucleus, where it can bind to heat shock elements (HSEs) to regulate transcription (Shamovsky et al. 2006). In addition, eEF1A1 associates with RNA polymerase II and the 3' UTR of heat shock protein 70 (HSP70) mRNA helping to stabilise and facilitate its transport from the nucleus to ribosomes (Vera et al. 2014). However, these roles are not shared by the eEF1A2 isoform (Vera et al. 2014).

Several other reports have also shown a role in nuclear transport. It has been demonstrated that eEF1A interacts with the nuclear export receptor exportin 5 (Exp5) in an aa-tRNA-dependent manner to support export of tRNAs (Bohnsack et al. 2002; Calado et al. 2002). This role may not be distinct from its role in translation, representing efficient channelling of translation components (Mateyak & Kinzy 2010). However, more recently eEF1A1 was demonstrated to form part of the nuclear export machinery (Khacho et al. 2008).

In addition, the two isoforms have opposing roles in apoptosis: eEF1A1 promotes apoptotic cell death in response to cell stress, while eEF1A2 promotes cell survival. Several studies have demonstrated that eEF1A1 is involved in the cytotoxic response to oxidative stress (Chen et al. 2000), ER stress and lipotoxicity (Borradaile et al. 2006). On the other hand, in response to stress cells expressing eEF1A2 have been shown to downregulate eEF1A2 and express eEF1A1 (Ruest et al. 2002). Furthermore, eEF1A2 directly interacts with Prx-1 to protect cells from stress-induced apoptosis (Chang & Wang 2007).

5.1.2.3.2 Roles in Disease

In addition to their non-canonical roles, each isoform has been implicated in various diseases, notably viral infections and cancer. Recently, the eEF1A2 isoform has been linked with intellectual disability, autism, and epilepsy.

The eEF1A proteins, particularly eEF1A1, are implicated in many stages of viral infection (reviewed by Li et al. 2013) and this is particularly apparent in the HIV-1 life cycle (Abbas et al. 2015). In addition, the eEF1A1 isoform has been shown to directly interact with the hepatitis B virus protein X (HBx), blocking its ability to bundle filamentous actin (Lin et al. 2012).

Upregulation of eEF1A2, and *EEF1A2* gene amplifications, are commonly found in many cancers, particularly in tissues which normally do not express eEF1A2; it is also associated with lung and pancreatic cancers (Abbas et al. 2015). Additionally, 60% of primary breast tumours, 30% of primary ovarian cancers and 43% of hepatocellular carcinomas (HCCs) have eEF1A2 overexpression (Abbas et al. 2015). It is believed eEF1A2 elicits its oncogenic effect through signalling pathways including Akt kinase and phosphoinositide 3-kinase (PI3K), resulting in cytoskeletal remodelling, proliferation, survival and invasion (Li et al. 2010; Pecorari et al. 2009; Amiri et al. 2007).

More recently it has emerged that mutations in *EEF1A2* leading to single amino acid changes in eEF1A2 can lead to autism, intractable epilepsy and intellectual disability (Nakajima et al. 2015; Veeramah et al. 2013; de Ligt et al. 2012). To date three different amino acid substitutions, found in four patients, have been published: Gly70Ser (Veeramah et al. 2013; de Ligt et al. 2012), Asp252His and Glu122Lys (Nakajima et al. 2015). All three of these residues are evolutionarily conserved and are found in, or close to, the regions involved in binding of eEF1B and GTP/GDP. It seems likely that these mutations are affecting translation but the precise underlying mechanisms remain to be elucidated.

5.1.3 Generation of *EEF1A1* Knockout Cell Lines

It is still unclear to what extent these additional roles (distinct from translation) overlap between the isoforms. Another important question is why the developmental switch happens, why some tissues replace one isoform with an almost identical isoform. This may be a consequence of the moonlighting functions of each isoform, which may be deleterious in some cell types.

I aimed to investigate these issues further by developing the tools with which I could start to answer these questions. Currently, the appropriate tools with which to untangle the different and potentially overlapping roles of the two isoforms are lacking; no cell lines expressing only the eEF1A2 isoform, are available. However, the availability of SSNs for precise genome editing, has meant that this should now be achievable.

5.2 Aims

Use the CRISPR/Cas9 system to knockout the function of *EEF1A1* in cell lines that express both isoforms in order to study the function of the eEF1A2 inform in isolation, to more fully understand its diverse roles in the cell.

5.3 Results

5.3.1 Expression of *EEF1A1* and *EEF1A2* in Immortalised Cell Lines

To create a cell line only expressing the eEF1A2 isoform, I needed to identify an appropriate starting line that expresses both isoforms. I selected a panel of immortalised human lines: RPE-1, HCT116, A549, HEK293FT, and HeLa cells. Three of these are cancer lines (A549, HCT116, and HeLa) and so have a good chance of expressing the eEF1A2 isoform. In order to quantify the expression of each isoform, I carried out qPCR analysis on these lines. The expression of each isoform was normalised to a panel of five reference genes, which were found to be most stable (as discussed in Section 3.2.1).

Figure 5.3A shows that *EEF1A1* is expressed at fairly comparable levels across all lines tested, except in HeLa cells in which it shows higher expression. On the other hand, *EEF1A2* displays more variability in its expression across the cell lines tested (Figure 5.3B), with highest expression seen in HCT116, A549 and HEK293FT cells.

Unfortunately, due to the lack of eEF1A2-specific antibodies, I was unable to investigate protein expression in each line, so I had to rely on the qPCR expression data. As HCT116 and HEK293FT cells demonstrated the highest *EEF1A2* expression, I chose to use these cell lines for all subsequent experiments.

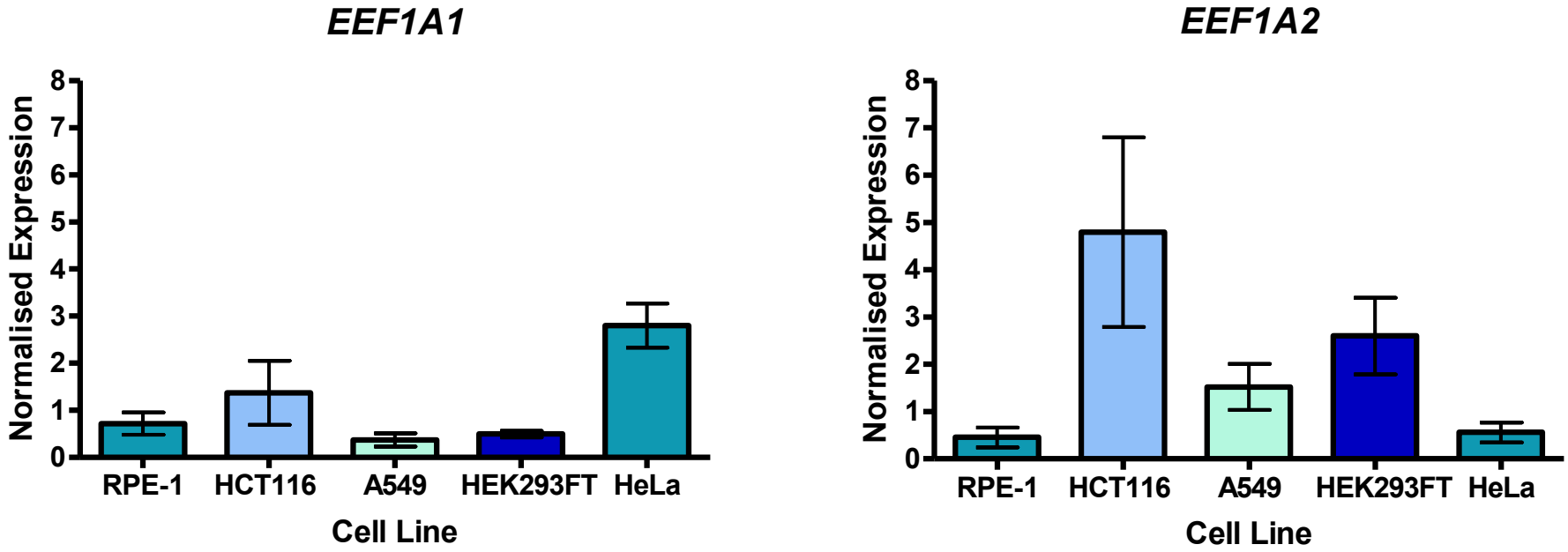


Figure 5.3 qPCR expression profiles for *EEF1A1* and *EEF1A2* in a panel of immortalised cell lines. Expression for each gene was normalised to a panel of five reference genes (*GAPDH*, *SDHA*, *UBC*, *TOP1* and *ATP5B*). Each bar represents the mean of three independent experiments. Error bars indicate S.E.M.

5.4 Targeting *EEF1A1* with CRISPR/Cas9

In order to knockout the function of eEF1A1 in cells, I planned to disrupt *EEF1A1* using CRISPR/Cas9 technology.

5.4.1 Identification of CRISPR/Cas9 Target Sites

Identification of unique target sites within *EEF1A1* is complicated by the presence of several retropseudogenes in the human genome. In order to improve the specificity of the sgRNAs, I chose to target the intronic sequences in the 5' portion of the *EEF1A1* gene, close to the intron-exon boundaries. I intended to use the double nickase CRISPR/Cas9 system to minimise the off-target activity, consequently, I identified four intronic sgRNA target sites, which could be used as single guides or as two nickase pairs, located upstream of exon three and downstream of exon four (Figure 5.4).

I constructed these sgRNAs using the Zhang laboratory method (Ran, P D Hsu, et al. 2013). I constructed these using both pX459 and pX462 plasmids, so that I could use them as either individual sgRNAs or as nickase pairs, respectively. I validated the activity of all individual guides and the two nickase pairs in HEK293FT cells using the SURVEYOR assay (data in Chapter 3), and was able to demonstrate robust activity.

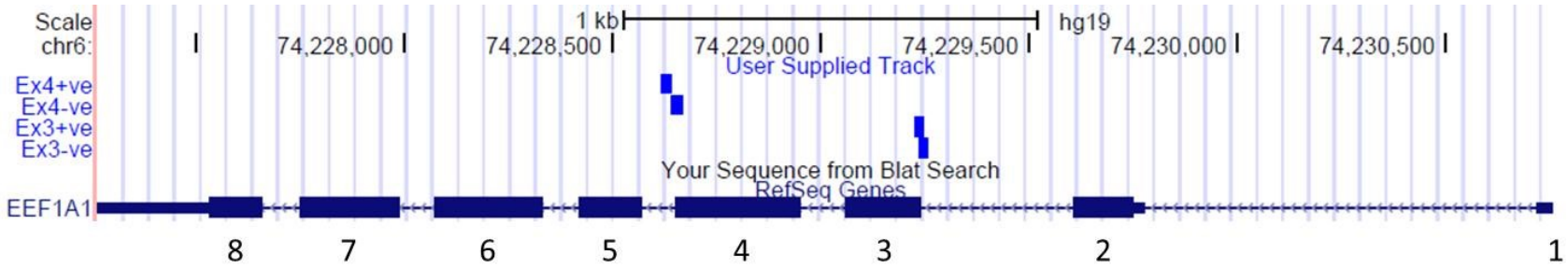


Figure 5.4 Genomic location of sgRNA target sites. The position of each sgRNA is indicated by a blue box. The sgRNAs are named after the closest exon and the DNA strand on which their target appears. The individual sgRNAs were constructed as both individual sgRNAs (pX459 backbone) and paired nickases (pX462 backbone). Exon numbers are indicated. Figure generated using UCSC genome browser, assembly GRCh37/hg19.

5.4.2 Generation of *EEF1A1* Knockout Cell Lines

I initially targeted *EEF1A1* in HEK293FT cells as I had confirmed the activity of each sgRNA in this line and the qPCR data indicated expression of the eEF1A2 isoform. Following targeting with the CRISPR/Cas9 system, both individual sgRNAs and nickase pairs, I isolated clonal lines using a FACS machine to sort individual cells into wells of a 96-well plate. This resulted in three surviving lines derived from sgRNA_EEF1A1_Ex3+ve, sgRNA_EEF1A1_Ex4+ve and sgRNA_EEF1A1_Ex4+ve/sgRNA_EEF1A1_Ex4-ve nickase pair transfections. Only one line (targeted with sgRNA_EEF1A1_Ex4+ve) showed evidence of indels at the target site following direct sequencing of the target region (Figure 5.5).

Figure 5.5 shows the sequence degeneration appearing close to the predicted cleavage site, 3-bp upstream of the PAM site (red dashed line). The chromatogram shows two distinct traces emerging at the site of cleavage. Based on sequence analysis, it appears that a deletion event has occurred.

I next used HCT116 cells, as they have a near normal karyotype with a modal chromosome number of 45, unlike HEK293FT cells whose modal chromosome number is 64. This would make subsequent studies of eEF1A2 function more meaningful as they would be in a near normal genetic background.

I used individual sgRNAs with wild type Cas9 for all subsequent experiments (unless otherwise stated) as several studies had indicated that off target activity was less common than first believed (Wang et al. 2015; Cho et al. 2014; Mandal et al. 2014; Smith et al. 2014; Suzuki et al. 2014; Veres et al. 2014).

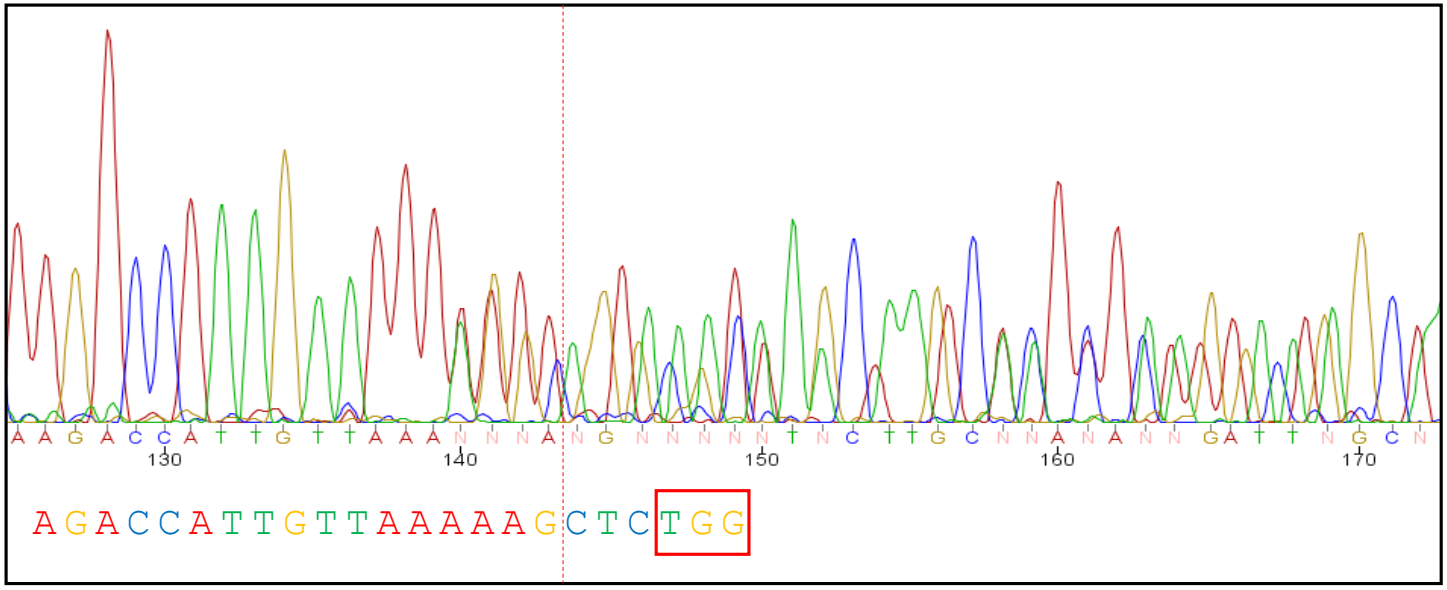


Figure 5.5 Chromatogram of an *EEF1A1* mutant clonal line. Clone was derived from the transfection of HEK293FT cells with sgRNA_EEF1A1_Ex4+ve. Red dashed line indicates predicted cleavage site. The sgRNA sequence is depicted below; PAM is indicated by red box.

I generated clonal lines, following transfection with all four individual sgRNAs in combination with wild type Cas9 and selection with puromycin, by FACS, resulting in 32 clonal lines. However, following sequencing, none of the 32 lines showed any evidence of indels at the target sites.

5.4.2.1 Targeting Both *EEF1A1* and *EEF1A2*

In order to assess the activity of the CRISPR/Cas9 system in HCT116 cells, I targeted both *EEF1A1* and *EEF1A*, comparing the ability of each to tolerate indels. In order to target *EEF1A2*, I used a nickase pair of sgRNAs, which were a kind gift from Faith Davies. This pair of sgRNAs target exon three of the *EEF1A2* gene (Figure 5.6) and have been shown to be active in cells (data not shown). To target *EEF1A1* I used sgRNA_EEF1A1_Ex3-ve and sgRNA_EEF1A1_Ex4-ve, as these had the least predicted off-target sites. Following transfection of HCT116 cells, puromycin selection and clone isolation by FACS, I was able to generate 14 clonal lines, 6 in which I had targeted *EEF1A1,* and 8 in which I had targeted *EEF1A2*.

Amplification of the *EEF1A2* target region indicated the presence of a deletion in one clone of approximately 300-bp (Figure 5.7A). Direct sequencing of the PCR products showed evidence of indels at the target site, indicated by the presence of multiple peaks in the chromatograms (Figure 5.7B). However, many of the sequences were too distorted to accurately interpret and due to technical issues, I was unable to clone and sequence the individual amplicons to reveal the nature of these mutations. In contrast, direct sequencing of PCR amplicons revealed no indels for any of the *EEF1A1* targeted clones.

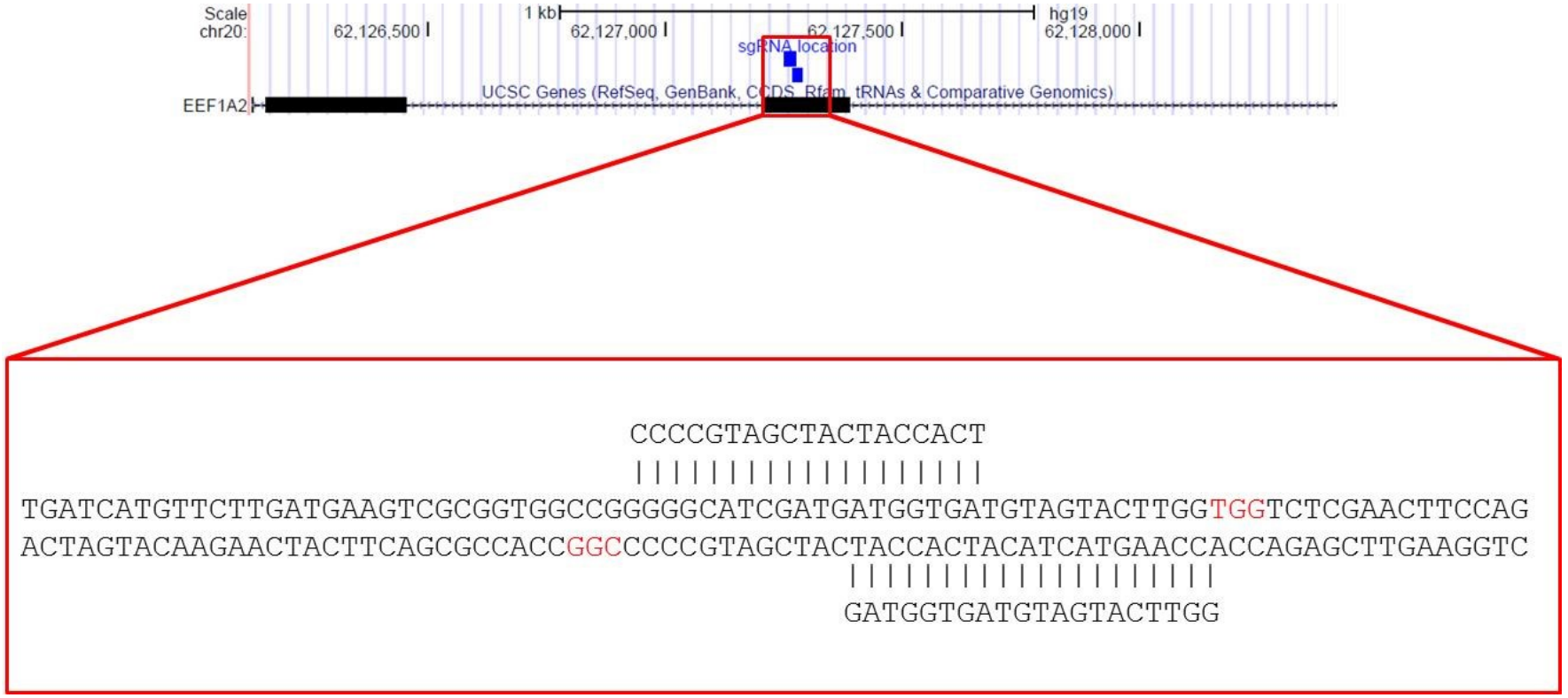


Figure 5.6 Genomic location of *EEF1A2*-targeting sgRNAs. Location of each sgRNA is indicated by the blue boxes. Sequenced for each sgRNA are given in the lower panel PAM is coloured red. Upper panel was generated using UCSC genome browser, assembly GRCh37/hg19.

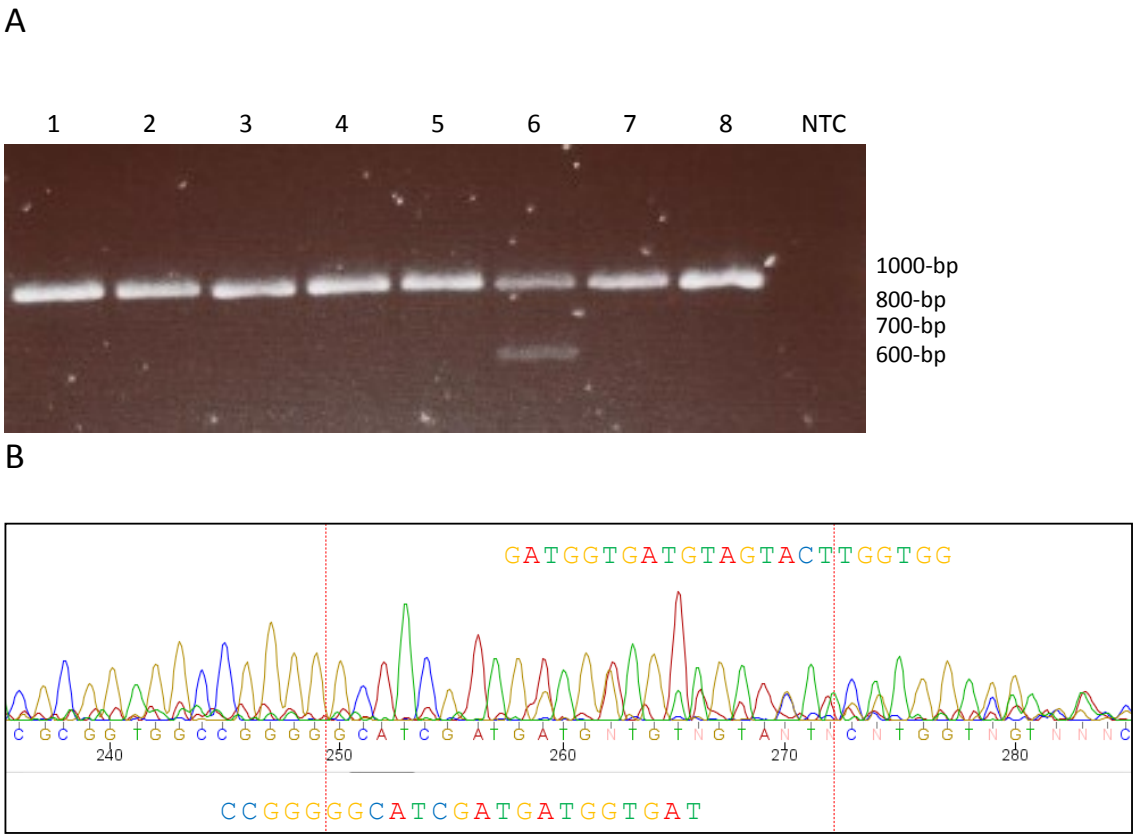


Figure 5.7 *EEF1A2*-targeted clonal lines. A) PCR amplification of target region. Clone 6 has an additional band suggestive of a deletion of approximately 300-bp. B) Example chromatogram showing degeneration of sequence at target site; trace from clone 7. Dotted red lines indicate predicted cleavage sites of the paired nickases. sgRNA target sites are indicated above and below.

5.4.2.2 Targeting Endogenous *EEF1A1* in the Presence of Exogenous *EEF1A1* and *EEF1A2*

As *EEF1A1* was proving difficult to disrupt, I wanted to see if I could target the endogenous *EEF1A1* in the presence of exogenous *EEF1A1* or *EEF1A2* (provided by cDNA expression plasmids). The *EEF1A1*-specific sgRNAs target intronic sequences and therefore would only target the endogenous gene and not the cDNA, so should have allowed me to mutate the endogenous locus whilst providing cells with their eEF1A needs. I also decided to target both HCT116 and HEK293FT cells, as HEK293FT cells had proven permissible to CRISPR/Cas9-induced indel formation at the *EEF1A1* locus and so would maximise the chances of producing a mutant line.

Before generating the clonal lines, I wanted to demonstrate that the cDNAs would express correctly in both cell lines. I therefore carried out a transient transfection with each construct, then a western blot to indicate protein expression.

Figure 5.8 demonstrates that cells expressed proteins of expected size following transfection with both the *EEF1A1* and *EEF1A2* cDNA expression plasmids, indicating correct expression of both isoforms.

I used only two gsRNAs, one targeting each intron; I chose sgRNA_EEF1A1_Ex3-ve and sgRNA_EEF1A1_Ex4+ve, as their target sites were entirely intronic and so should not have impaired exogenous eEF1A1 function. In addition, I had already demonstrated the ability of sgRNA_EEF1A1_Ex4+ve to induce indels in HEK293FT cells (Figure 5.5).

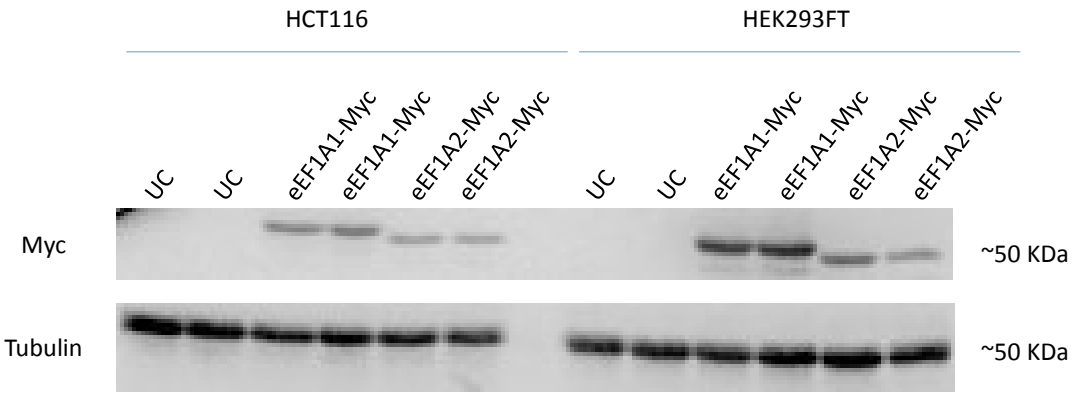


Figure 5.8 Western blot following transient transfection of *EEF1A* cDNAs. Constructs are tagged with a Myc tag, allowing for detection using an anti-Myc antibody. Both constructs are expressed in both HEK293FT and HCT116 cells.

Table 5.1 Clonal Lines. Clonal lines were generated by isolating single cells following transfection of either HEK293FT or HCT116 cells with sgRNA_EEF1A1_EX3-ve and sgRNA_EEF1A1_Ex4+ve in combination with wild type Cas9. Some cells were co-transfected with exogenous expression plasmids encoding either *EEF1A1* or *EEF1A2*. A total of 36 clonal lines were generated.

Clone	Cell Line	sgRNA	cDNA
1	HEK293FT	Ex3-/Ex4+	
2	HEK293FT	Ex3-/Ex4+	
3	HEK293FT	Ex3-/Ex4+	
4	HEK293FT	Ex3-/Ex4+	
5	HEK293FT	Ex3-/Ex4+	
6	HEK293FT	Ex3-/Ex4+	
7	HEK293FT	Ex3-/Ex4+	
8	HEK293FT	Ex3-/Ex4+	
9	HEK293FT	Ex3-/Ex4+	pcDNA3.1:hEEF1A1-Myc
10	HEK293FT	Ex3-/Ex4+	pcDNA3.1:hEEF1A1-Myc
11	HEK293FT	Ex3-/Ex4+	pcDNA3.1:hEEF1A1-Myc
12	HEK293FT	Ex3-/Ex4+	pcDNA3.1:hEEF1A1-Myc
13	HEK293FT	Ex3-/Ex4+	pcDNA3.1:hEEF1A1-Myc
14	HEK293FT	Ex3-/Ex4+	pcDNA3.1:hEEF1A2-Myc
15	HEK293FT	Ex3-/Ex4+	pcDNA3.1:hEEF1A2-Myc
16	HEK293FT	Ex3-/Ex4+	pcDNA3.1:hEEF1A2-Myc
17	HEK293FT	Ex3-/Ex4+	pcDNA3.1:hEEF1A2-Myc
18	HCT116	Ex3-/Ex4+	
19	HCT116	Ex3-/Ex4+	
20	HCT116	Ex3-/Ex4+	
21	HCT116	Ex3-/Ex4+	
22	HCT116	Ex3-/Ex4+	
23	HCT116	Ex3-/Ex4+	pcDNA3.1:hEEF1A1-Myc
24	HCT116	Ex3-/Ex4+	pcDNA3.1:hEEF1A1-Myc
25	HCT116	Ex3-/Ex4+	pcDNA3.1:hEEF1A1-Myc
26	HCT116	Ex3-/Ex4+	pcDNA3.1:hEEF1A1-Myc
27	HCT116	Ex3-/Ex4+	pcDNA3.1:hEEF1A1-Myc
28	HCT116	Ex3-/Ex4+	pcDNA3.1:hEEF1A1-Myc
29	HCT116	Ex3-/Ex4+	pcDNA3.1:hEEF1A2-Myc
30	HCT116	Ex3-/Ex4+	pcDNA3.1:hEEF1A2-Myc
31	HCT116	Ex3-/Ex4+	pcDNA3.1:hEEF1A2-Myc
32	HCT116	Ex3-/Ex4+	pcDNA3.1:hEEF1A2-Myc
33	HCT116	Ex3-/Ex4+	pcDNA3.1:hEEF1A2-Myc
34	HCT116	Ex3-/Ex4+	pcDNA3.1:hEEF1A2-Myc
35	HCT116	Ex3-/Ex4+	pcDNA3.1:hEEF1A2-Myc
36	HCT116	Ex3-/Ex4+	pcDNA3.1:hEEF1A2-Myc

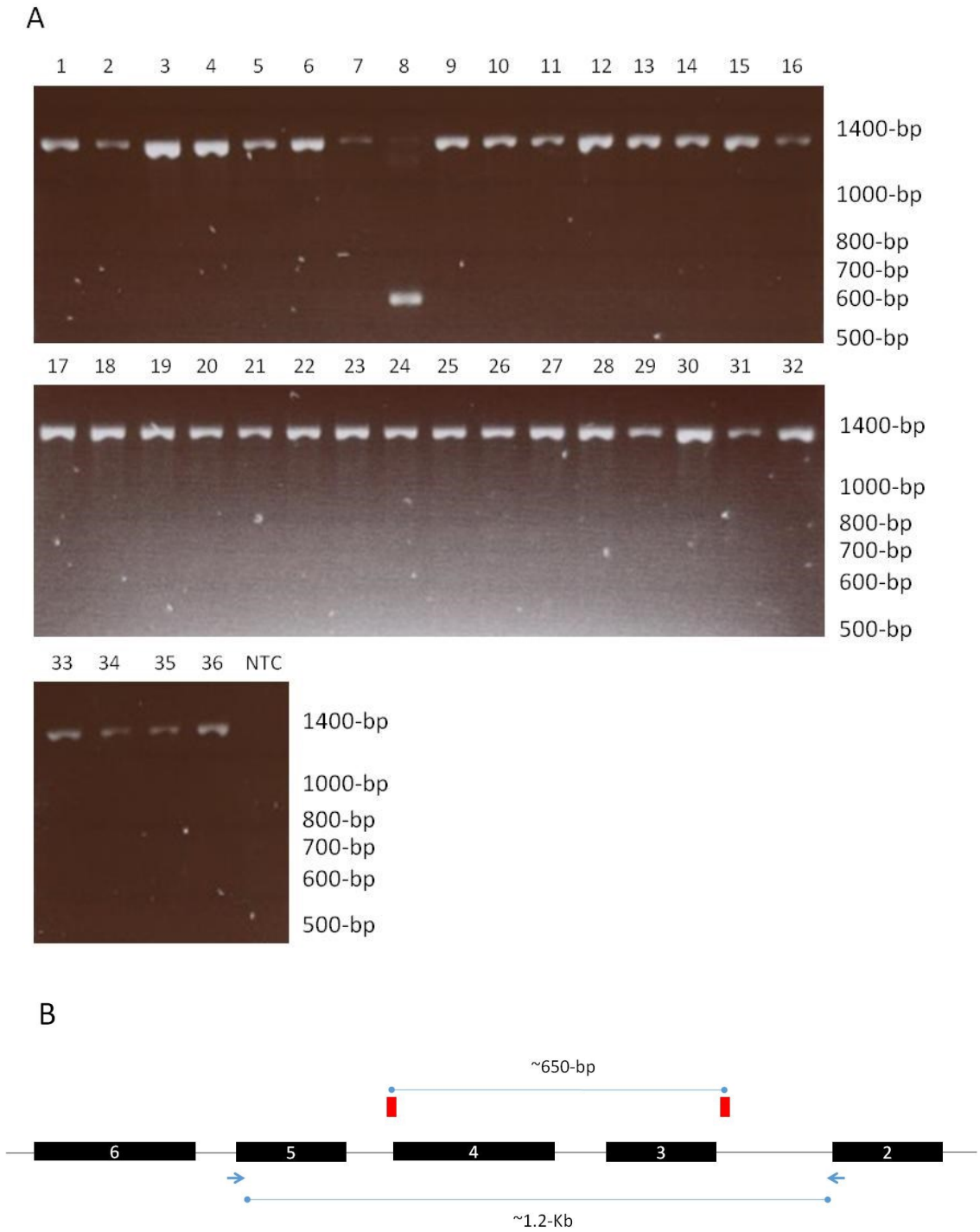


Figure 5.9 PCR analysis of *EEF1A1* target region. Primers surrounding both target sites were used to identify any large deletions between the two sites. A) Agarose gel electrophoreses of amplified target region. Clone 8 shows evidence of a large, approximately 650-bp deletion. Clone information is detailed in Table 5.1. NCT = no template control. B) Schematic of target region, sgRNA target sites are indicated by the red boxes. Primers used to amplify the target region are indicated by the blue arrows. Approximate sizes of PCR amplicons and potential deletion region are indicated.

I was able to generate 19 HCT116 clones and 17 HEK293FT clones (Table 5.1). Sequencing confirmed that none of the HCT116 lines had mutations in *EEF1A1*, while only one HEK293FT clonal line showed any evidence of *EEF1A1* mutation. In this case, PCR amplification of the target site indicated the presence of a large deletion equal to the loss of exons three and four (~650-bp) on one allele (Figure 5.9). Two additional bands were present for this clone (clone 8), indicating that three alleles are present in HEK293FT cells, two of which appear to have deletions in this clonal line, as indicated by the smaller bands on the agarose gel. Sequencing did not reveal the nature of the mutation on the remaining allele(s) and, due to time constraints, I was unable to investigate this clone further.

5.4.2.3 Characterisation of the *EEF1A1* mutant Line Containing an Intronic Deletion

As discussed in Section 5.3.2, I was able to generate a clonal line in HEK293FT cells, which carried an indel. In order to investigate this further I cloned the PCR product and sequenced the individual amplicons.

Figure 5.10 shows that this clone is a compound heterozygote, carrying a deletion of 8-pb on one allele and 12-bp on the other. Both deletions centre on the predicted cleavage site 3-bp upstream of the PAM (highlighted in red). This deletion occurs within intron four, 28-bp downstream of exon four. Two amplicons also demonstrate base substitutions near the cleavage site (Figure 5.10, red letters).

In order to determine any effect on the splicing of the eEF1A1 protein, I carried out a western blot alongside a wild type HEK293FT sample in order to observe any size difference.

	A G A C C A T T G T T A A A A A G C T C T G G G A A T G G C G A T T T C A T G C T T A C A C A A A
TOPO Clone 3	A G A C C A T T G T T A A A - - - - - - - - - - - - - T G G C G A T T T C A T G C T T A C A C A A A
TOPO Clone 6	A G A C C A T T G T T A A A - - - - - - - - - - - - - T G G C G A T T T C A T G C T T A C A C A A A
TOPO Clone 9	A G A C C A T T G T T A A A - - - - - - - - - - - - - T G G C G A T T T C A T G C T T A C A C A A A
TOPO Clone 10	A G A C C A T T G T T A A A - - - - - - - - - - - - - T G G C G A T T T C A T G C T T A C A C A A A
TOPO Clone 13	A G A C C A T T G T T A A A - - - - - - - - - - - - - T G G C G A T T T C A T G C T T A C A C A A A
TOPO Clone 14	A G A C C G T T G T T A A A - - - - - - - - - - - - - T G G C G A T T T C A T G C T T A C A C A A A
TOPO Clone 26	A G A C C A T T G T T A A A - - - - - - - - - - - - - T G G C G A T T T C A T G C T T A C A C A A A
TOPO Clone 7	A G A C C A T T G T T A A A A A A - - - - - - - - A A T G G C G A T T T C A T G C T T A C A C A A A
TOPO Clone 5	A G A C C A T T G T T A A A A A A - - - - - - - - A A T G G C G A T C T C A T G C T T A C A C A A A
TOPO Clone 16	A G A C C A T T G T T A A A A A A - - - - - - - - A A T G G C G A T T T C A T G C T T A C A C A A A

del 12
del 12
del 12
del 12
del 12
del 12, A>G
del 12
del 8
del 8, T>C
del 8

Figure 5.10 Sequence analysis of *EEF1A1* mutant line. Individual amplicons were sequences to reveal the extent of the mutation in the clonal line. Two deletions appear to be present, a 12-bp deletion, and an 8-bp deletion. Green boxes = sgRNA target; red boxes = PAM; grey boxes = no change; white box = indel; (-) = deletion; red letter = substitution; del = deletion; ins = insertion; > = substitution; WT = wild type.

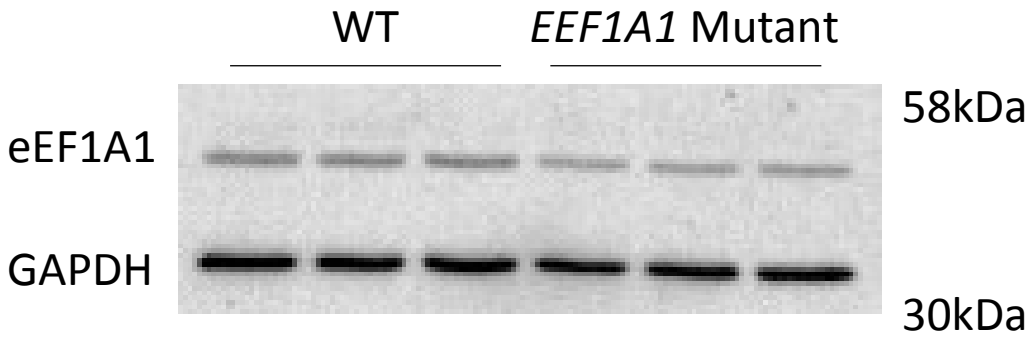


Figure 5.11 Western blot of *EEF1A1* mutant and wild type. There is no evidence of altered protein size in the mutant line, indicated by the presence of a band of the same size in both samples.

Figure 5.11 shows that there is no observable difference in protein size between the wild type and the mutant lines. However, it appears that expression in the mutant line is less than that of the wild type cells. In order to investigate this I quantified expression based on three independent experiments (normalised to GAPDH expression).

Figure 5.12 suggests there might be a lower level of protein in the mutant cell lines, but there is no significant difference between eEF1A1 protein expression in the mutant compared to the wild type ($p = 0.77$).

To investigate the effect of the mutation on transcription, I carried out a qPCR on both wild type and mutant cells.

Figure 5.13 shows that the mutant line has an approximately six-fold *increase* in the expression of *EEF1A1* at the transcriptional level compared to wild type ($p = 0.043$).

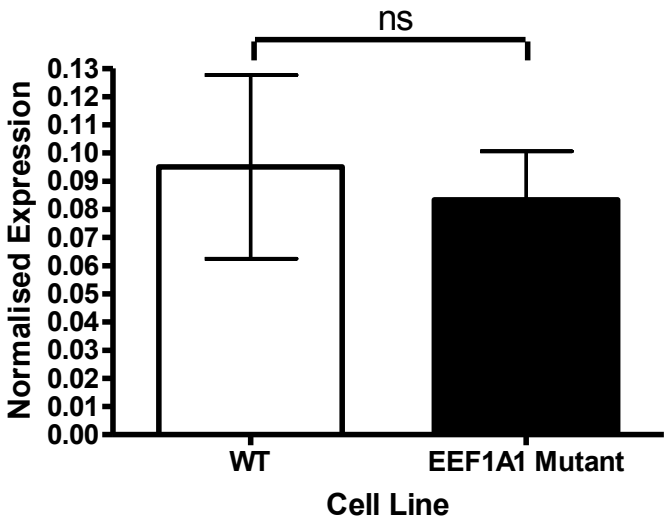


Figure 5.12 Protein expression analysis. Data represent the means of three independent western blots. Expression was normalised to the expression of the loading control, GAPDH. Error bars indicate S.E.M. Unpaired T-test was used to calculate the p-value. ns = not significant; WT = wild type.

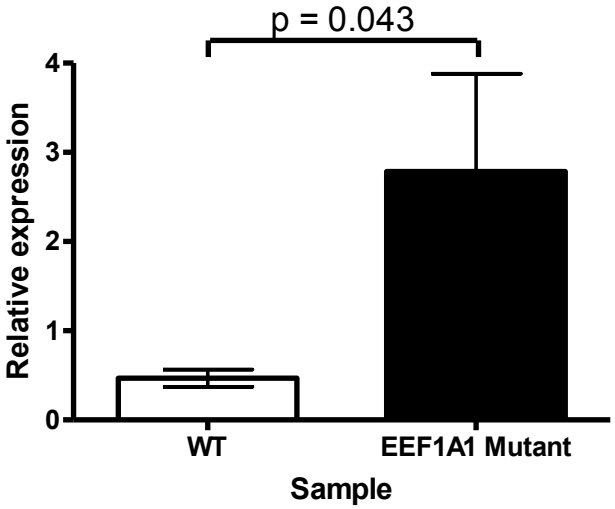


Figure 5.13 qPCR analysis of *EEF1A1* expression. Data represent the means of three independent experiments. Expression was normalised to the expression of five reference genes. Error bars indicate S.E.M. P-value calculated using unpaired T-test; WT = wild type.

5.4.2.3.1 Regulatory Elements

To investigate the increase in transcriptional activity observed in the mutant line compared to wild type cells, I looked at the regulatory data provided by the ENCODE project; I identified a DNase I hypersensitivity cluster which coincides with the site of the deletion (Figure 5.14). In addition, there are four transcription factor (TF) binding sites, which are potentially disrupted in the mutant line: chromodomain helicase DNA binding protein 1 (CHD1), polymerase (RNA) II (DNA directed) polypeptide A (POLR2A), metastasis associated 1 family, member 3 (MTA3), and signal transducer and activator of transcription 5A (STAT5A) (Figure 5.15).

I also used JASPAR software to predict if any TF binding sites were created by the deletion; 10 TF binding sites are potentially created, these are listed in Table 5.2.

Table 5.2 Predicted transcription factor binding sites in *EEF1A1* deletion mutant. Mutant sequences for each allele were used to predict the generation of TF binding sites using JASPAR.

Transcription Factor	Full Name
CEBPA	CCAAT/Enhancer Binding Protein (C/EBP), Alpha
CEBPB	CCAAT/Enhancer Binding Protein (C/EBP), Beta
FOXH1	Forkhead Box H1
THAP1	THAP Domain Containing, Apoptosis Associated Protein 1
NFIC	Nuclear Factor I/C (CCAAT-Binding Transcription Factor)
STAT3	Signal Transducer and Activator Of Transcription 3 (Acute-Phase Response Factor)
SRY	Sex Determining Region Y
SOX9	SRY (Sex Determining Region Y)-Box 9
YYI	YY1 Transcription Factor
SOX10	SRY (Sex Determining Region Y)-Box 10

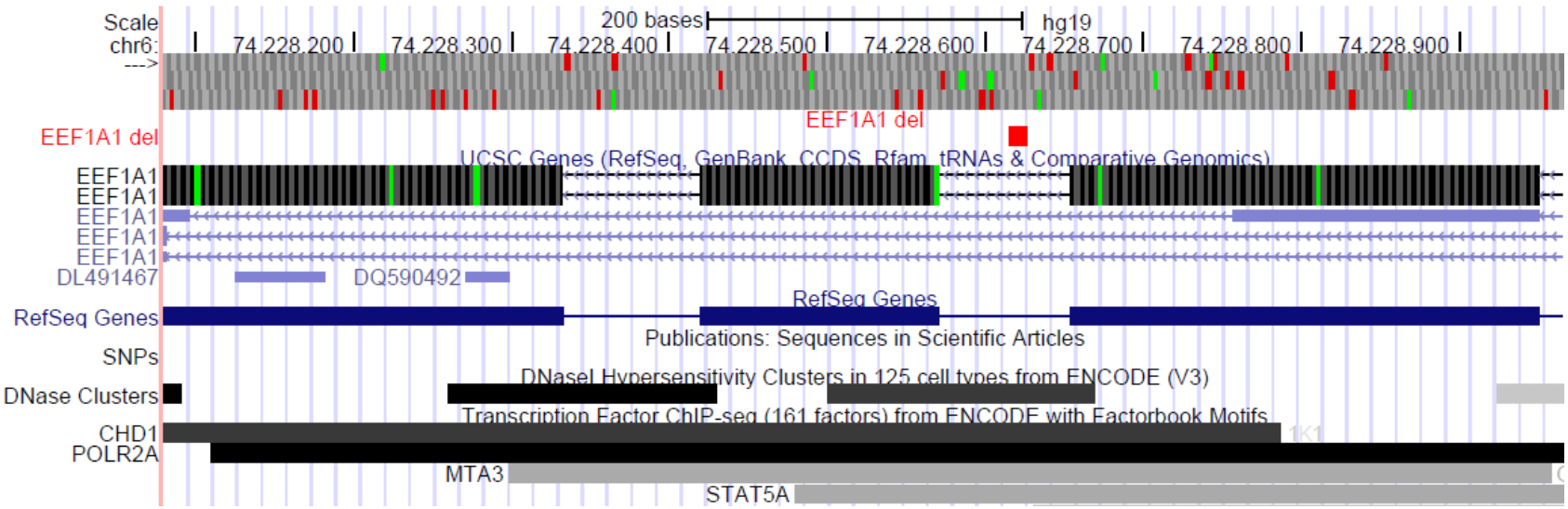


Figure 5.14 Regulatory elements at site of *EEF1A1* deletion. Deletion is indicated by the red box. Data from the ENCODE project, available through UCSC genome browser. The site of the deletion coincides with both a DNase I hypersensitivity cluster and four transcription factor binding sites. Figure generated using UCSC genome browser, assembly GRCh37/hg19.

5.5 Discussion

The CRISPR/Cas9 system has been used by many groups to knockout genes in different organisms and cell lines (Chu et al. 2015; Wettstein et al. 2015; Zhou et al. 2015; Han et al. 2014; Shalem et al. 2014b; Yang et al. 2014). However, the creation of an *EEF1A1* knockout cell line has proved challenging here. All rounds of targeting in HCT116 failed to generate any clonal lines carrying a mutation at the *EEF1A1* locus. While I have generated two mutant lines in HEK293FTs, one of these at least maintains a normal eEF1A1 protein level. It is tempting to speculate that eEF1A1 may be essential even in the presence of eEF1A2; however, much work remains to confirm this.

5.5.1 Endogenous Expression

One of the challenges in studying eEF1A isoforms is the lack of specific antibodies. Most of the eEF1A2-specific antibodies will in fact, also cross-react with eEF1A1. Consequently, in order to determine expression, I had to rely upon qPCR data, which may not necessarily reflect the protein expression within the cell. The qPCR data indicated that eEF1A2 is expressed most highly in HCT116, HEK293FT and A549 cells. This is consistent with other reports which have demonstrated expression at the transcriptional, and in some cases protein level, in these lines (Krastel et al. 2015; Kawamura et al. 2014; Kim et al. 2009).

5.5.2 Targeting *EEF1A1* and *EEF1A2*

Initial targeting of *EEF1A1* in HCT116 proved unsuccessful, with none of the 32 clonal lines showing any evidence of CRISPR/Cas9 activity. This could have been due to lack of CRISPR/Cas9 activity in this cell line. To investigate this further, I decided to target both *EEF1A1* and *EEF1A2*. As I was only able to generate 14 clonal lines, it is difficult to draw any firm conclusions, but in light of all the rounds of cloning it seems plausible to speculate that mutations in *EEF1A1* are less well tolerated than mutations in *EEF1A2,* at least in HCT116 cells. In total, I was able to generate 57 *EEF1A1*-targeted clonal lines and eight *EEF1A2* clonal lines in HCT116; none of these demonstrated indels at the *EEF1A1* locus. On the other hand, at least two of the eight lines had indels at the *EEF1A2* locus. Potentially more of the lines also carried *EEF1A2* mutations, but I was unable to confirm this due to technical

issues. While not conclusive, this suggests that mutations in *EEF1A2* may be better tolerated than *EEF1A1* mutations, though it will be necessary to fully characterise the nature of the mutations observed in *EEF1A2*. However, it is also possible that the *EEF1A2*-targeting sgRNA nickase pair is more active in HCT116 cells than the *EEF1A1*-targeting sgRNAs. This seems unlikely as I demonstrated robust activity of these sgRNAs in HEK293FT cells, so this would have to be due to locus-specific difference between the two cell lines. It is possible that for sgRNA__EEF1A1_Ex3-ve, activity is reduced by the presence of a SNP at the target site; rs2073466 coincides with position one of the sgRNA target site, though, it is not known if HCT116 carries this SNP. However, it is unlikely that this would impede activity, as it has been shown that sgRNAs can tolerate mismatches in the 5' end (Fu et al. 2013; Hsu et al. 2013; Mali, Aach, et al. 2013; Pattanayak et al. 2013). In addition, this does not affect the activity of sgRNA_EEF1A1_Ex4+ve, so cannot entirely account for the lack of indels observed.

It seems plausible to hypothesise that mutations in *EEF1A1*, leading to alteration in expression and protein function, are deleterious to cell health, even in the presence of eEF1A2, and so clones die; surviving clones are those which do not carry any mutations at this locus. As mutations in *EEF1A2* were observed, it is possible that this isoform is not essential in the presence of eEF1A1, though further investigation is required to confirm the nature of these mutations and the consequences on protein level.

5.5.3 Targeting Endogenous *EEF1A1* in the Presence of Exogenous eEF1A

I was unable to generate *EEF1A1* mutant lines in HCT116 cells, even in the presence of exogenous eEF1A. In addition, I was unable to generate mutant lines from HEK293FT cells in the presence of exogenous eEF1A. This is potentially due to the loss of the exogenous expression, due to the transient nature of plasmid expression.

One HEK293FT clone does carry a large deletion in the *EEF1A1* gene on one allele, equivalent in size to loss of exons three and four, and potentially a smaller deletion on a second allele. It seems likely that a deletion of exons three and four would lead to a knock out of the gene function, though further investigation is required to

confirm this. However, as this is a monoallelic disruption, it is unlikely to cause a total loss of eEF1A1 function.

Other studies have shown disruptions in *EEF1A1* in the presence of eEF1A2. In their study Potts and colleagues (2015) used the eEF1A1 inhibitor didemnin B which stabilises the interaction of aa-tRNA and eEF1A1, preventing translation. In this study, they demonstrated that resistant cells were enriched for eEF1A2 expression, suggesting that eEF1A2 may be able to account for loss of eEF1A1 function. However, it is not known if didemnin B also inhibits other functions of eEF1A1 outside its role in translation, which do not require its aa-tRNA binding ability. It is possible that in these cells eEF1A1 is still able to carry out other non-canonical functions and that eEF1A2 only accounts for loss of eEF1A1-mediated translation. On the other hand, Borradaile et al. (2006) generated CHO cells carrying a disruption in the *EEF1A1* gene. These cells also express the eEF1A2 isoform, so one could speculate that loss of eEF1A1 is accounted for by eEF1A2. However, these cells still expressed eEF1A1, though at a greatly reduced level. It is therefore possible that eEF1A2 provided the translation capacity whilst the, albeit reduced, eEF1A1 carried out other essential functions in these cells. In addition, Krastel and colleagues (2015) demonstrated that the cytotoxic effect of nanocystin A from myxobacteria, were due to its inhibition of eEF1A1 and not eEF1A2 in HCT116 cells. This suggests that eEF1A1 is essential in these cells, as inhibition, even in the presence of eEF1A2, leads to cell death.

5.5.4 *EEF1A1* Mutant Line

I was able to generate a clonal line from HEK293FT cells with an intronic deletion in *EEF1A1*. Sequencing analysis confirmed the presence of two mutant alleles – 8-bp and 12-bp deletions occurring in intron four. In addition to the deletions, two amplicons demonstrated base substitutions; these were likely the result of PCR errors rather than erroneous DSB repair. It seems possible that there are three *EEF1A1* alleles present in HEK293FT cells (Section 5.3.2.2), though I only observed two in this clone; further investigation is required to rule out the presence of a wild type allele.

It was possible that these deletions may have knocked out sequences important for correct splicing. However, I was able to demonstrate that protein size in the mutant line is the same as that in the wild type line, indicating that correct splicing of exon four is not inhibited in the mutant line. This does not rule out the possibility of impaired splicing efficiency. In addition, this line showed no significant difference in protein expression, consistent with the hypothesis that eEF1A1 protein is essential even in the presence of eEF1A2. However, it displayed significantly increased levels of *EEF1A1* transcription, ~6-fold greater than the wild type cells.

The location of the deletion in this line coincides with a DNase I hypersensitivity site. This suggests that this region is likely to be important in the regulation of expression, indeed the ENCODE data suggest that there are four TF binding sites within this region: CHD1, POLR2A, MTA3 and STAT5A. While these data are experimentally generated, this work was not conducted in HEK293FT cells and so may not indicate TF binding in this cell line. MTA3 is the only one of these TFs associated with transcriptional repression (Brüning et al. 2014) as such seems a likely candidate to explain the increased expression observed; loss of binding could lead to a loss of repression and therefore an increase in transcription. In addition, using JASPAR software, I identified 10 potential TF binding sites, created by the mutations. All 10 TFs are associated with promoting transcription so could potentially be involved in the upregulation observed in the mutant, though any interact would need to be validated experimentally.

Other intronic deletions have been shown to cause increased gene expression. McCready and colleagues (1997) found that deletions in the first intron of murine CD4, also coinciding with DNase I hypersensitivity sites, led to an increase in expression possibly due to loss of negative regulatory elements.

The difference in expression seen at the level of transcription and translation is consistent with reports that demonstrate the up regulation of the protein in response to cellular stress. Protein levels are increased at the level of translation, but not at the level of transcription following stress (Chen et al. 2000), probably to allow for rapid increases in protein level when required. This suggests that regulation of

transcription and translation may be uncoupled and that fine-tuning of the protein level can occur by regulation at the translational rather than transcriptional level.

5.6 Future Perspectives

Due to the time constraints of this project, I have been unable to fully characterise the mutant lines I generated. Initial investigation of the line bearing the intronic deletion has shown its altered transcription of the *EEF1A1* gene. It would be worth investigating this further and identifying the specific factor(s) involved.

For the line bearing the predicted exonic deletion, it would be important to fully characterise the nature of this deletion. Any effect this has on both transcription and translation could then be investigated. Should protein level be reduced, it would be possible to investigate changes in cell proliferation and growth, protein synthesis, and response to cellular stress such as heat shock or oxidative stress, which would be suggestive of isoform-specific differences.

In parallel, it would be appropriate to investigate the *EEF1A2* mutant lines, generated in HCT116 cells. Firstly, it would be necessary to characterise fully the nature of the mutations, by cloning and sequencing individual amplicons. These mutant lines would provide a useful tool to study the consequences of eEF1A2 disruption; abnormalities in functions associated with one or other isoform in these cells could be attributed to loss, or reduced function of eEF1A2.

A recent study has indicated that small molecules can be used to modulate the activity of the CRISPR/Cas9 system, favouring either HR or NHEJ depending on the desired outcome (Yu et al. 2015). In this study, the authors demonstrated that azidothymidine (AZT), a drug used in HIV/AIDS treatment, could be used to promote NHEJ repair at the expense of the HR repair pathway, thus increasing the knockout efficiency. It would be worth investigating if the use of AZT could aid in the generation of *EEF1A1* knockout cells.

An alternative method to generate an *EEF1A1* knockout would be to introduce LoxP or flippase recognition target (FRT) sites both sides of a critical exon. The introduction of the LoxP or FRT sites could be facilitated by the sgRNAs I have constructed here. LoxP and FRT sites are short sequences (34-bp), which are

recognised by Cre and Flp recombinases, respectively; if two sites are placed flanking an exon, expression of the recombinase will lead to a deletion of the exon between the two recognition sites. This method has been demonstrated by Chen et al. (2015) who introduced biallelic FTR sites flanking an exon of *PAX6* using sgRNAs targeting either side of the exon and a plasmid donor. A similar method could be used to target *EEF1A1* using the sgRNAs I have generated here. This would demonstrate if eEF1A1 truly is essential even in the presence of eEF1A2; cells would die upon expression of the recombinase if eEF1A1 expression is essential.

Chapter 6 Concluding Remarks

6.1 Targeting Genomic Loci Using SSNs

The main aim of this project was to develop SSNs to target four genomic loci, which I had chosen as candidates to support transgene integration and long-term expression. I chose these loci based on evidence suggesting that they are widely expressed. In the case of the AAVS1 and hROSA26 loci, their potential to support transgene integration had been demonstrated, though I aimed to investigate this further. In addition, I intended to develop a drug-selectable transgene integration system at the *IMPDH2* locus, and an *EEF1A1* knockout cell line, in order to probe the biology of eEF1A1 and the closely related eEF1A2. In light of this, I designed and constructed a range of TALENs and individual sgRNAs to target each locus. I validated the ability of each to induce indels at their target site, and demonstrated activity in all cases using the SURVEYOR assay. Due to the highlighted incidence of off-target activity during the early development of the CRISPR/Cas9 system, I also constructed sgRNA nickase pairs for three of my chosen sites; for most of these, I was able to demonstrate robust activity. However, those targeting *IMPDH2* failed to show robust activity; this is likely due to the non-optimal orientation of the sgRNA pairs.

6.1.1 Transgene Integration

The four genomic loci I chose are all widely expressed, and I was able to confirm transcriptional activity at all loci in a range of immortalised human cell lines. This indicates that each has an open chromatin structure, which is more supportive of transgene expression. The AAVS1 locus was used as a control locus: SSNs had previously been developed to target this region, and others had demonstrated its potential to support transgene expression. I was able to confirm the ability of these SSNs to induce indels at their target sites and I confirmed that AAVS1 could support integration of transgenes. In doing so, I demonstrated the ability of both TALENs and CRISPR/Cas9 to drive the integration of large transgenes. This gave me confidence that I could achieve this at the other loci.

Building on this, I intended to integrate the *mCherry* gene at the hROSA26 locus, stimulated by both TALENs and CRISPR/Cas9. However, due to unanticipated cloning issues, I did not achieve this in the time available, though I have developed the tools with which this locus may be probed in future. The potential of this site to

support transgene integration, and subsequent expression, even after stem cell differentiation, has been indicated (Irion et al. 2007). This site also appears to be evolutionarily conserved (at least in mammals), with homologues recently identified in rat (Kobayashi et al. 2012) and pig (Li et al. 2014), in addition to mouse, in which it was first found (Friedrich & Soriano 1991). In all cases, it appears this locus can support transgene integration and robust expression. It is likely that the human ROSA26 locus will prove to be a good site for transgene integration, though further work will be required to validate this.

Whilst selection of permissive sites for transgene integration can help to reduce the chances of transgene silencing, it is important to take into consideration other factors both within the local environment and the transgene cassette, which can influence transgene expression. The mouse *Rosa26* locus exemplifies this: transgene expression levels demonstrate significant orientation-dependence. Chen et al. (2011) showed that some promoters, most notably EF1α, demonstrated a significant difference in transgene expression when integrated in the same orientation as *Rosa26* compared to integration in the opposite orientation to *Rosa26.* In all cases, integration in the same orientation as *Rosa26* resulted in greatest transgene expression. However, in their study, Strathdee and colleagues (2006) demonstrated that transgenes integrated in the opposite orientation to *Rosa26* showed higher levels of expression. They also showed that deletion of upstream sequences reduced transgene expression when integrated in the same orientation as *Rosa26*, suggesting that there is transcriptional interference from the endogenous *Rosa26* promoter. The disparity between these studies is likely due to the presence of an insulator sequence in the transgene cassette in the Chen et al. (2011) study, which would have prevented any interference from the endogenous promoter. Taken together, these studies suggest a potential role for orientation in transgene integration and indicate the importance of careful cassette design, to prevent interference from endogenous signals.

Due to time limitations, I could not investigate the ability of all of the chosen loci to support transgene integration, though I have developed the tools with which this can be achieved in the future.

6.2 Development of a Drug-Selectable Transgene Integration System

The development of a drug-selectable transgene integration system was hampered by the unexpected leakiness of MPA treatment. The ability of MPA to inhibit IMPDH is well documented, but this does not account for other cellular pathways that are capable of bypassing IMPDH in fulfilling the cell's guanine nucleotide needs, notably the salvage pathway, resulting in cellular survival despite MPA treatment. While the proposed system proved to be impractical to implement, I was subsequently able to demonstrate that *IMPDH2* could support transgene integration and robust long-term expression. However, this work was conducted in cultured cells, therefore further work would be required to investigate this *in vivo*.

Of the sites I had chosen, *IMPDH2* demonstrated the highest level of transcription across all cell lines tested. This is indicative of an open chromatin structure, permissive to gene expression, indicating its potential to support transgene expression. It is clear that local environmental factors can influence transgene expression, as discussed above; it would be useful investigating further the link between levels of endogenous expression and the ability to support transgene expression.

It is clear from the data presented here that the CRISPR/Cas9 system can efficiently target transgene integration: at *IMPDH2* 68.1% of clones carried a correctly integrated transgene. In addition, *IMPDH2* can support robust expression, with 37.5% of the correctly integrated clones demonstrating high-level mCherry expression. However, it will be necessary to fully characterise each of these clonal lines, particularly why some lines with correct integration, have low, or no transgene expression.

6.3 Targeting *EEF1A1*

The data presented here potentially indicate that eEF1A1 is essential even in the presence of eEF1A2, as evidenced by the inability to generate mutations resulting in a loss of protein expression. This remains to be confirmed for the mutant line carrying a potential exonic deletion. In addition, I was able to generate several clonal

lines bearing *EEF1A2* mutations, which could suggest that the eEF1A2 isoform is less essential: full characterisation of these mutant lines will be necessary to confirm this. This supports the hypothesis that each isoform has a unique profile of moonlighting functions in the cell, such that eEF1A2 is unable to account for a loss of eEF1A1. Should this be the case, then this could account for the switch in expression that occurs during early postnatal development. The cells that normally express the eEF1A2 isoform tend to be terminally differentiated, and so have different requirements than rapidly dividing cells, which could be one of the reasons for the isoform switch.

Though unable to create an *EEF1A1* knockout line, I was able to generate a clonal line carrying a biallelic deletion within intron four. In addition, I generated a clonal line with a potential monoallelic deletion of exons three and four. Interestingly, for the intronic deletion mutant, I was able to show that the total eEF1A1 protein was *reduced* compared to control cells, though this did not reach significance ($p = 0.77$), while the level of transcription was significantly *increased* compared to controls ($p = 0.043$). The reasons for this were not clear, though it seems likely that binding of regulatory factors was altered by the deletion. It is also possible that the deletion causes reduced mRNA stability, such that increased transcription is required to maintain normal protein level. Further investigation is required to elucidate the underlying mechanisms.

6.4 Future Directions

6.4.1 Validation of *IMPDH2* and hROSA26

In this thesis, I have demonstrated the potential of *IMPDH2* to act as a site for integration and long-term expression of transgenes. More work is required to validate fully the robustness of this locus for this purpose. Some clonal lines with correct transgene integration did not demonstrate mCherry expression; it will be necessary to investigate if this is due to integration errors or transgene silencing. If the latter, it would be useful to investigate how common this is in subsequent rounds of targeting. It would also be useful to investigate the silencing mechanisms involved, so that they may be avoided through improved cassette design.

The data here indicate that choice of sgRNA may influence the efficiency of integration and subsequent expression of the transgene at *IMPDH2*. It would be useful to investigate this further, to identify the underlying reasons behind this, which could influence the design of more efficient sgRNAs in the future.

Though I was unable to characterise hROSA26 as a site for transgene integration, others have indicated its potential (Irion et al. 2007), and so it would be beneficial to investigate this further. Firstly, it will be necessary to investigate if the SSNs I have developed here can drive transgene integration and if these can be expressed long-term from this locus. As mentioned above, the mouse *Rosa26* site shows significant orientation-dependent expression of transgenes. It would be worth investigating if this is also true for the human gene.

IMPDH2 shows the highest level of transcription of all the sites I investigated. It would be useful to compare expression of integrated transgenes to endogenous expression, and to investigate this relationship, which could be useful in the prediction of other sites useful for transgene integration.

Once optimal targeting conditions have been found for each site, it would be necessary to integrate a transgene of biological or therapeutic interest, in order to demonstrate the utility of each site for relevant applications.

6.4.2 Targeting *EEF1A1*

In light of the data presented here, it seems likely that the use of a more sophisticated strategy may be necessary in order to knockout *EEF1A1*. This could involve conditional knockout mediated by LoxP/Cre or FRT/Flp systems. This would help to determine if eEF1A1 is essential even in the presence of eEF1A2, as induction of Cre or Flp, resulting in *EEF1A1* knockout, would cause cell death if this is the case. Another option would be to use SSNs to direct changes in *EEF1A1* to make it more *EEF1A2*-like. This would help to determine which regions are important to the activity of each isoform. However, based on the evidence gathered here, this is likely to be a significant technical challenge.

6.4.3 Investigating *EEF1A1* Mutant Lines

Initial investigations here demonstrated an interesting disparity between protein and mRNA levels in one of the mutant lines. It would be worth investigating the underlying causes of this, which could uncover novel regulatory pathways.

In addition, it will be necessary to characterise the second clonal line. This line appears to carry a large deletion on one allele, which corresponds to a loss of exons three and four. It is likely that this will have some effect on expression, though the extent of this remains unknown.

6.4.4 Improving SSN Efficiency

In their recent study, Yu and colleagues (2015) identified small molecule drugs, which could be used to bias the outcome of CRISPR/Cas9 cleavage towards either NHEJ or HR repair pathways, increasing gene disruption or repair, respectively. It would be useful to investigate the effect of these chemicals with the SSNs I have developed here, to increase the efficiency of either transgene integration or gene knockout.

6.5 Comparison of the SSN Technologies

The development of the CRISPR/Cas9 system has made genome editing more accessible to a greater number of scientists due to the simplicity of design and ease of construction. While this system is undoubtedly a powerful tool, the other available SSN systems (namely TALENs, ZFNs and to some extend meganucleases) all have their own merits. In most comparisons, TALENs were more highly active than ZFNs (Chen, Oikonomou, et al. 2013), and due to their modular nature and relative ease of design, the TALEN system rapidly grew in popularity. Studies directly comparing CRISPR/Cas9 and TALEN technologies are few, but do suggest that the CRISPR/Cas9 is more active in most cases (Mandal et al. 2014; Ding et al. 2013). In addition, I observed greater activity for CRISPR/Cas9 in direct comparisons at both AAVS1 and hROAS26.

ZFNs showed significant context dependence: the ability of each ZnF to bind its target could be influenced by neighbouring ZnFs (Carroll 2011). This has not been observed for TALENs and due to the nature of target recognition, is not an issue for

CRISPR/Cas9. In addition, the proprietary nature of the system meant that it was difficult to generate highly active ZFNs to target loci of interest.

The identification of TALEN target sites is limited by several factors. The requirement of FokI to dimerise means that the paired binding sites must be situated 14–20-bp apart (Sanjana et al. 2012). While this potentially limits the number of targetable sequences in the genome, it also helps to increase the fidelity of the system as such sites are less likely to occur outside of the target locus. Indeed, TALENs often show very little evidence of off-target activity (Wang et al. 2015; Smith et al. 2014; Veres et al. 2014). Due to a cryptic signal within the N-terminal, TALEN target sites must begin with a T nucleotide; together with the necessity of the 14-20-bp spacer region, this can reduce the number of targetable sequences, especially compared to CRISPR/Cas9. In addition, several other factors have been suggested to improve the activity of the TALENs, which may further reduce the number of potential binding sites (Cermak et al. 2011). However, the number of repeats in the DNA-binding domain can vary greatly: naturally occurring TALEs can contain up to 33.5 repeats and as few as 6.5 have been shown to be active (Boch et al. 2009). This greatly increases the repertoire of targetable sites.

On the other hand, the CRISPR/Cas9 system is only limited by the need for a PAM site, in the form 5'-NGG for *S. pyogenes* Cas9, immediately 3' of the 20-bp target sequence. However, Cas9 molecules from other bacterial species, with different PAM requirements have been isolated, including *Neisseria meningitides* (5'-NNNNGATT), *Streptococcus thermophilus* (5'-NNAGAAW), *Treponema denticola* (5'-NAAAAC) and *Staphylococcus aureus* (5'-NNGRR[T]) (Ran et al. 2015; Cong et al. 2013; Esvelt et al. 2013; Hou et al. 2013). In addition, the specificity of *S. pyogenes* Cas9 has been reengineered to allow alternative PAM recognition (Kleinstiver et al. 2015). Together, these could be used to increase the repertoire of targetable sequences within the genome.

One of the advantages of the CRISPR/Cas9 system is its compact nature, which makes it particularly amenable to multiplexing. Several studies have shown the benefits of this approach, including in the induction of targeted gross chromosomal rearrangements, namely inversions and translocations (Choi & Meyerson 2014). This

has also allowed for the development of large-scale, genome-wide libraries for genome wide recessive screens (Bassett et al. 2015; Koike-Yusa et al. 2013; Shalem et al. 2014a; Y. Zhou et al. 2014).

One of the practical advantages of the CRISPR/Cas9 system is the ease of construction compared with the other SSN system. A single cloning step is required to produce a new sgRNA, compared to the more labour intensive construction of TALENs. In my experience, the construction of sgRNAs is generally more efficient, resulting in the majority of isolated plasmids carrying the correct insert. On the other hand, due the more complex nature of the TALEN construction process, I found that many of the resulting plasmids had misincorporation of the monomer repeats. However, newer more streamlined TALEN construction techniques have been developed, which allow for high-throughput production, such as the FLASH (fast ligation-based automatable solid-phase high-throughput) system (Reyon et al. 2012), though it is likely that the simplicity of CRISPR/Cas9 construction will lead to its replacement of TALENs in many laboratories.

6.6 Off-Target Activity

An issue, raised early in the generation of the CRISPR/Cas9 system, was its propensity for off-target activity. It was demonstrated that each sgRNA could tolerate as many as five mismatches between the RNA and the target site, particularly in the 5' portion of the sgRNA, distal to the PAM site (Fu et al. 2013; Hsu et al. 2013; Mali, Aach, et al. 2013; Pattanayak et al. 2013). Later studies also demonstrated the ability of 5'-NAG to be recognized by Cas9, rather than the expected 5'-NGG PAM site (Anders et al. 2014). In response to these issues, a double nickase strategy was devised, making use of the D10A mutant Cas9 (Cas9n) in combination with two sgRNAs targeting opposing DNA strands (Mali, Aach, et al. 2013). In light of these issues, I chose to implement the double nickase strategy for my target sites. However, due to issues regarding sgRNA orientation discussed earlier, the nickase pairs targeting *IMPDH2* failed to show detectable activity. More recently, groups have also developed a FokI-Cas9 fusion, which in combination with paired sgRNAs, causes DSBs in an analogous manner to TALENs and ZFNs (Guilinger, Thompson, et al. 2014; Tsai et al. 2014). Unlike the paired nickase

strategy, FokI will not cut unless it dimerises and so activity is more restricted: binding sites must be located so that FokI is able to dimerise, further increasing the fidelity of the system (Tsai et al. 2014). However, data from whole genome sequencing experiments suggest that the off-target profile for the CRISPR/Cas9 system may be less of an issue than first believed, particularly when carefully designed sgRNAs are used (Wang et al. 2015; Cho et al. 2014; Mandal et al. 2014; Smith et al. 2014; Suzuki et al. 2014; Veres et al. 2014). In addition, use of truncated sgRNAs, removing the 5' portion, which is most prone to tolerate mismatches, can further increase the sgRNA fidelity (Fu et al. 2014). It would be useful to investigate the true burden of off-target activity for each of my sgRNAs (and TALENs), though this is beyond the scope of this thesis.

Off-target activity is not unique to the CRISPR/Cas9 system. Though off-target activity does not appear to be a problem for TALENs (Wang et al. 2015; Smith et al. 2014; Veres et al. 2014), it was a significant concern for ZFNs, with some showing significant cytotoxicity due to the burden of DSBs (Alwin et al. 2005; Porteus & Baltimore 2003; Bibikova et al. 2002). Methods to reduce the off-target cleavage by ZFNs were developed, but in many cases led to reduced on-target activity (Söllü et al. 2010; Miller et al. 2007; Szczepek et al. 2007). On the other hand, meganucleases show probably the greatest specificity in target recognition of all the SSNs (Harrison et al. 2014), but they have become less frequently used due to the difficulty in reengineering their target recognition.

6.7 Wider Applications of SSN Technology

In this work, I have focused on the use of SSN technology to drive transgene integration, which can be particularly useful for biotechnology, for example the development of stable lentiviral producer cell lines or therapeutic protein production systems. However, this technology has myriad potential applications, in basic research, gene therapy, and the wider scientific community.

In the study of gene function, it could be argued that the need for permissive loci is negated by the availability of SSNs, which allow scientists to target almost any position within the genome, thus allowing direct modifications of genes of interest. While this is useful in many applications, the ability to introduce transgenes site-

specifically into cells is still useful for biotechnology applications, basic research and has potential for future gene therapy. The ability to express any gene of interest or genes carrying specific mutations is essential to the study of gene function and consequences of patient specific mutations. Advances in stem cell technology, particularly iPSCs, means that patient-specific mutations can be studied in isogenic backgrounds and in disease-relevant cell types, which would otherwise be impossible to study, such as neurones. The hROSA26 locus has already demonstrated its potential to maintain transgene expression after stem cell differentiation (Irion et al. 2007). In addition, this technology allows genes to be studied in cells in which they would not normally be expressed.

These tools have huge potential for biotechnology. The range of therapeutic proteins is ever increasing, particularly monoclonal antibodies. The production of these proteins has often been carried out in bacteria, due to the low cost culturing and ease of engineering (Kamionka 2011). However, this can lead to problems with incorrect folding, processing and post translational modifications, as such mammalian cell systems have become increasingly utilised (Wurm 2004). However, transgenes can be subject to silencing. To overcome this, sites which have been demonstrated to be less prone to silencing, such as the four potential sites I have chosen here, can be exploited.

Other systems to create stable transgenic cell lines are available, but these generally rely upon recombinase mediated transgene insertion such as the Flp-In™ and Jump-In™ platforms available from Life technologies in a range of cell backgrounds (Cat. No's A14148, A14150, A15007, A15008, R78007, R76007, R75807, R75207, R75007, R76207, and R76107). While these are useful tools, they only allow for insertions at single sites, so would not be applicable for the generation of stable lentiviral producer cell lines, or in other instances where two or more independent transgene integration events are desirable.

6.7.1 Gene Therapy

This technology provides new and exciting possibilities for gene therapy, both in their ability to promote specific HR events, for gene correction or gene addition, and in their ability to create specific DNA breaks, for both gene knockout and repair of

trinucleotide repeat expansions. Early studies in yeast have demonstrated the ability of TALENs to specifically target and delete trinucleotide repeat tracts (Richard 2015). Subsequent studies in Friedreich's ataxia patient-derived cells, demonstrated that ZFNs could be used to specifically target the trinucleotide repeat region and led to a significant increase in frataxin expression, which persisted through reprogramming into iPSCs and differentiation into neurones (Li et al. 2015). In addition, the development of iPSCs and the means to specifically alter them with SSNs opens up new potentials for personalised medicine. For example, Yusa and colleagues (2011) demonstrated that targeted correction of the α1-antitrypsin gene (*A1AT*) could be achieved by ZFNs in patient-derived iPSCs and that this restored function of A1AT following differentiation into liver cells. Other groups have also shown the potential of this strategy for diseases such as SCID-X1. In their study, Menon et al. (2015) demonstrated that TALENs could be used to correct mutations in the gene encoding IL-2Rγ in patient-derived iPSCs, which could then be successfully differentiated to generate mature natural killer cells and T-cell precursors, demonstrating the potential for autologous cell therapy. The ability to manipulate patient cells *ex vivo* also holds great potential for many other diseases. For example Osborn et al. (2013) illustrated the potential of TALENs to drive HR repair of mutations in the *COL7A1* gene in fibroblasts derived from recessive dystrophic epidermolysis bullosa (RDEB) patients. These demonstrated normal protein expression in a mouse model following transplantation of iPSCs.

In addition, the ability to selectively disrupt gene function has allowed researchers to target the *CCR5* gene, which encodes the major co-receptor for HIV-1. A mutation in *CCR5* has been identified in individuals who are resistant to HIV-1 infection. The mutation, termed Δ32 (a deletion of 32-bp) results in a severely truncated protein which is not detected at the cell surface, and so HIV-1 cannot gain entry (Liu et al. 1996). The first human clinical trial used ZFNs to knockout *CCR5* function, which would mimic the effect of the natural Δ32 mutation (Tebas et al. 2014). This trial demonstrated the potential and safety of this strategy: most patients showed a reduction in the detectable levels of circulating HIV-1 (Tebas et al. 2014). In addition, some studies have also focused on eradication of integrated virus using TALENs and CRISPR/Cas9 (Hu et al. 2014; Ebina et al. 2015; Strong et al. 2015).

The ability to integrate a gene into a locus that will support robust long-term expression also has great potential for future gene therapy. The use of SSN-mediated targeted integration is likely to be less efficient than integration following retroviral delivery. However, the ability to predefine a specific integration site, should be much safer than random integration, which in rare cases can lead to activation of oncogenes (Hacein-Bey-Abina et al. 2008). One of the disadvantages of the CRISPR/Cas9 system is the large size of the Cas9 protein (4-Kb coding sequence) which could reduce the efficiency of its delivery and choice of vectors for gene therapy. However, alternative Cas9 proteins from other bacterial species have been identified, which have a smaller coding sequence, but equivalent cleavage efficiencies (Ran et al. 2015). One of the major problems with SSNs, particularly CRISPR/Cas9, for gene therapy is their potential for off-target activity. However, as more is learnt about the system, better designs will allow for highly specific sgRNAs, which have already been demonstrated for some loci (Veres et al. 2014; Smith et al. 2014).

Diseases such as CF, which result from mutations in a single gene leading to loss of a functional protein, are excellent candidates for gene therapy. In the case of CF, provision of the *CFTR* gene can lead to an improvement of symptoms, as has been demonstrated in early clinical trials (Alton et al. 2015). SSNs could be used to repair the mutations in *CFTR;* indeed studies using ZFNs and CRISPR/Cas9 have demonstrated the potential of this (Schwank et al. 2013; Lee et al. 2012). However, replacement of the gene, by integration into a validated locus, could be used as a one-size-fits-all solution. The advantage over a patient specific approach is that this only requires the testing and validation of a single set of SSNs and integration cassette, rather than individual testing of mutation-specific targeting strategies, which would each require trails for safety and efficacy. Such an approach could be applicable to many diseases, which result from single gene defects.

6.7.2 Modification of Human Embryos

The public acceptance of gene therapy rests on the belief that it will be used only to treat diseases, not for enhancement, and that these alterations will be in somatic cells, so will not be passed on to future generations. However, the development of SSNs,

particularly the CRISPR/Cas9 system, means that the possibility of specifically manipulating human embryos is now more feasible. In their landmark study, Liang and colleagues (2015) demonstrated that high frequency gene editing is achievable in human embryos. The study investigated the ability of CRISPR/Cas9 to drive repair of mutations in the β-globin gene (*HBB*), associated with β-thalassemia, in non-viable triponuclear human embryos. While the authors demonstrated efficient cleavage of the target site, repair using the donor was relatively inefficient. They also discovered that endogenous homologous templates, notably the δ-globin gene (*HBD*), could compete with the exogenous repair templates. This phenomenon has also been shown in mouse embryos (Wu et al. 2013). In addition, many of the edited embryos were mosaic, though whether this is linked to their triponuclear nature is unknown. Interestingly, the fidelity of sgRNA binding appears less stringent in the embryos than in HEK293T cells; none of the predicted off target sites were cleaved in the cell line, but two of the top seven sites were cleaved in the embryos. In addition, further off target cleavage was confirmed by whole-exome sequencing. While the technology is far from being specific enough for the precise manipulation required, we are closer than ever to this possibility. This study has caused substantial controversy, raising many questions about the ethics of embryo manipulation. One of the major reservations of such an approach is the unknown side effects of editing within the embryo: do the potential benefits outweigh the risks? One could argue that this type of manipulation is unnecessary; such a technique will require screening of manipulated embryos to ensure the fidelity of the indented correction, therefore one could screen unedited embryos to identify those not carrying the disease variant, negating the need for gene editing and its associated risks. It seems that the technology is moving faster than the legislation and a more in depth discussion within the scientific community, and society in general, is required to identify a realistic framework for future human genome modification.

6.7.3 Modification of Disease-Carrying Organisms

Another use of CRISPR/Cas9 technology is to engineer ‘gene-drives’ in mosquitoes that would render them resistant to *Plasmodium* and make them less fertile, which could help to reduce the incidence of malaria. However, we must also be aware of the potential consequences of such work, which could render the mosquito extinct,

having knock-on effects within the ecosystem, specifically for bats and other insect-eating species. In addition, it is vital to ensure that these modifications cannot be passed to other species, which could lead to unpredictable environmental consequences.

6.7.4 De-extinction

One of the more unusual applications to emerge from this technology is the de-extinction movement, which aims to restore extinct species such as the passenger pigeon and the woolly mammoth. However, de-extinction is not a new idea, for example, researches have been able to resurrect the genome of the 1918 influenza virus, but the technology to facilitate it has become more widely available and affordable. While this area of research is controversial, there are some potential benefits. One of the major proponents is George Church, whose group were one of the first to develop the CRISPR/Cas9 system, and is keen to promote the benefits of this idea. It is believed that bringing back ancient diversity could help to reverse radical changes in ecosystems, and may even help to slow the effects of global warming (Church 2013). In addition to bringing back extinct species, this technique could help existing species by reintroducing lost diversity in those that have become too inbred, such as the Tasmanian devil.

6.8 Closing Remarks

In this thesis, I have demonstrated some of the potential of SSN technology and identified a novel site for transgene integration, *IMPDH2*. Genome engineering is a rapidly evolving field and it is clear that SSNs are useful tools, which will become increasingly integral to basic research, biotechnology, and gene therapy. The future holds many exciting opportunities and it is likely that increasingly sophisticated approaches to gene editing will be developed.

Chapter 7 References

Abbas, W., Kumar, A. & Herbein, G., 2015. The eEF1A Proteins: At the Crossroads of Oncogenesis, Apoptosis, and Viral Infections. *Frontiers in oncology*, 5, p.75. Available at: http://journal.frontiersin.org/article/10.3389/fonc.2015.00075/abstract [Accessed June 5, 2015].

Ackermann, M. et al., 2014. Promoter and lineage independent anti-silencing activity of the A2 ubiquitous chromatin opening element for optimized human pluripotent stem cell-based gene therapy. *Biomaterials*, 35(5), pp.1531–42. Available at: http://www.sciencedirect.com/science/article/pii/S014296121301377X [Accessed January 18, 2016].

Aherne, A. et al., 2004. On the molecular pathology of neurodegeneration in IMPDH1-based retinitis pigmentosa. *Human molecular genetics*, 13(6), pp.641–50. Available at: http://hmg.oxfordjournals.org/content/13/6/641.full [Accessed May 26, 2015].

Aiuti, A. et al., 2013. Lentiviral hematopoietic stem cell gene therapy in patients with Wiskott-Aldrich syndrome. *Science (New York, N.Y.)*, 341(6148), p.1233151. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4375961&tool=pmcentrez&rendertype=abstract [Accessed June 25, 2015].

Allison, A.C. et al., 1977. The role of de novo purine synthesis in lymphocyte transformation. *Ciba Foundation symposium*, (48), pp.207–24. Available at: http://www.ncbi.nlm.nih.gov/pubmed/415850 [Accessed July 20, 2015].

Alton, E.W.F.W. et al., 2015. Repeated nebulisation of non-viral CFTR gene therapy in patients with cystic fibrosis: a randomised, double-blind, placebo-controlled, phase 2b trial. *The Lancet Respiratory Medicine*. Available at: http://www.thelancet.com/article/S2213260015002453/fulltext [Accessed July 6, 2015].

Alwin, S. et al., 2005. Custom zinc-finger nucleases for use in human cells. *Molecular therapy : the journal of the American Society of Gene Therapy*, 12(4), pp.610–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16039907 [Accessed August 7, 2015].

Amiri, A. et al., 2007. eEF1A2 activates Akt and stimulates Akt-dependent actin remodeling, invasion and migration. *Oncogene*, 26(21), pp.3027–40. Available at: http://dx.doi.org/10.1038/sj.onc.1210101 [Accessed August 2, 2015].

Ammar, I., Izsvák, Z. & Ivics, Z., 2012. The Sleeping Beauty Transposon Toolbox. In Y. Bigot, ed. *Mobile Genetic Elements*. Humana Press, pp. 229–240. Available at: http://dx.doi.org/10.1007/978-1-61779-603-6_13.

Anders, C. et al., 2014. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature*.

Antoniou, M. et al., 2003. Transgenes encompassing dual-promoter CpG islands from the human TBP and HNRPA2B1 loci are resistant to heterochromatin-mediated silencing. *Genomics*, 82(3), pp.269–279. Available at: http://www.sciencedirect.com/science/article/pii/S0888754303001071 [Accessed January 18, 2016].

El Ashkar, S. et al., 2014. BET-independent MLV-based Vectors Target Away From Promoters and Regulatory Elements. *Molecular therapy. Nucleic acids*, 3, p.e179. Available at: http://dx.doi.org/10.1038/mtna.2014.33 [Accessed July 28, 2015].

Asokan, A., Schaffer, D. V & Samulski, R.J., 2012. The AAV vector toolkit: poised at the clinical crossroads. *Molecular therapy : the journal of the American Society of Gene Therapy*, 20(4), pp.699–708. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3321598&tool=pmcentrez&rendertype=abstract [Accessed January 15, 2016].

Auer, T.O. et al., 2014. Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. *Genome research*, 24(1), pp.142–53. Available at: http://genome.cshlp.org/content/24/1/142.short

[Accessed February 8, 2016].

Bannister, A.J. & Kouzarides, T., 2011. Regulation of chromatin by histone modifications. *Cell research*, 21(3), pp.381–95. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3193420&tool=pmcentrez&rendertype=abstract [Accessed July 9, 2014].

Barski, A. et al., 2007. High-resolution profiling of histone methylations in the human genome. *Cell*, 129(4), pp.823–37. Available at: http://www.sciencedirect.com/science/article/pii/S0092867407006009 [Accessed July 9, 2014].

Bassett, A.R., Kong, L. & Liu, J.-L., 2015. A genome-wide CRISPR library for high-throughput genetic screening in Drosophila cells. *Journal of Genetics and Genomics*, 42(6), pp.301–309. Available at: http://www.sciencedirect.com/science/article/pii/S1673852715000594 [Accessed April 27, 2015].

Bateman, J.R., Lee, A.M. & Wu, C., 2006. Site-specific transformation of Drosophila via phiC31 integrase-mediated cassette exchange. *Genetics*, 173(2), pp.769–77. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1526508&tool=pmcentrez&rendertype=abstract [Accessed January 8, 2016].

Beil, J. et al., 2012. Is BAC transgenesis obsolete? State of the art in the era of designer nucleases. *Journal of biomedicine & biotechnology*, 2012, p.308414. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3413995&tool=pmcentrez&rendertype=abstract [Accessed January 18, 2016].

Belteki, G. et al., 2003. Site-specific cassette exchange and germline transmission with mouse ES cells expressing phiC31 integrase. *Nature biotechnology*, 21(3), pp.321–4. Available at: http://dx.doi.org/10.1038/nbt787 [Accessed January 19, 2016].

Bibikova, M. et al., 2002. Targeted chromosomal cleavage and mutagenesis in

Drosophila using zinc-finger nucleases. *Genetics*, 161(3), pp.1169–75. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1462166&tool=pmcentrez&rendertype=abstract [Accessed August 7, 2015].

Biffi, A. et al., 2013. Lentiviral hematopoietic stem cell gene therapy benefits metachromatic leukodystrophy. *Science (New York, N.Y.)*, 341(6148), p.1233158. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23845948 [Accessed July 4, 2015].

Bischoff, C. et al., 2000. The human elongation factor 1 A-2 gene (EEF1A2): complete sequence and characterization of gene structure and promoter activity. *Genomics*, 68(1), pp.63–70. Available at: http://www.sciencedirect.com/science/article/pii/S0888754300962712 [Accessed August 1, 2015].

Boch, J. et al., 2009. Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science*, 326(5959), pp.1509–1512. Available at: http://www.sciencemag.org/content/326/5959/1509.abstract.

Boch, J. & Bonas, U., 2010. Xanthomonas AvrBs3 Family-Type III Effectors: Discovery and Function. In *Annual Review of Phytopathology, Vol 48*. Palo Alto: Annual Reviews, pp. 419–436. Available at: <Go to ISI>://WOS:000282062400020.

Bochtler, M., 2012. Structural basis of the TAL effector–DNA interaction. *The Journal of Biological Chemistry*, 393(10), pp.1055–1066. Available at: http://www.degruyter.com/view/j/bchm.2012.393.issue-10/hsz-2012-0164/hsz-2012-0164.xml.

Bogdanove, A.J., Schornack, S. & Lahaye, T., 2010. TAL effectors: finding plant genes for disease and defense. *Current Opinion in Plant Biology*, 13(4), pp.394–401. Available at: http://www.sciencedirect.com/science/article/pii/S1369526610000531.

Bohnsack, M.T. et al., 2002. Exp5 exports eEF1A via tRNA from nuclei and

synergizes with other transport pathways to confine translation to the cytoplasm. *The EMBO Journal*, 21(22), pp.6205–6215. Available at: http://emboj.embopress.org/content/21/22/6205.abstract [Accessed August 3, 2015].

Borradaile, N.M. et al., 2006. A Critical Role for Eukaryotic Elongation Factor 1A-1 in Lipotoxic Cell Death. *Molecular Biology of the Cell*, 17(2), pp.770–778. Available at: http://www.molbiolcell.org/content/17/2/770.abstract.

Böttcher, R. et al., 2014. Efficient chromosomal gene modification with CRISPR/cas9 and PCR-based homologous recombination donors in cultured Drosophila cells. *Nucleic acids research*, 42(11), p.e89. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4066747&tool=pmcentrez&rendertype=abstract [Accessed July 18, 2015].

Braun-Sand, S.B. & Peetz, M., 2010. Inosine monophosphate dehydrogenase as a target for antiviral, anticancer, antimicrobial and immunosuppressive therapeutics. *Future medicinal chemistry*, 2(1), pp.81–92. Available at: http://www.future-science.com/doi/abs/10.4155/fmc.09.147?url_ver=Z39.88-2003&rfr_id=ori%3Arid%3Acrossref.org&rfr_dat=cr_pub%3Dpubmed& [Accessed July 17, 2015].

Broussau, S. et al., 2008. Inducible packaging cells for large-scale production of lentiviral vectors in serum-free suspension culture. *Molecular therapy : the journal of the American Society of Gene Therapy*, 16(3), pp.500–7. Available at: http://dx.doi.org/10.1038/sj.mt.6300383 [Accessed July 24, 2015].

Brüning, A. et al., 2014. Function and regulation of MTA1 and MTA3 in malignancies of the female reproductive system. *Cancer metastasis reviews*, 33(4), pp.943–51. Available at: http://link.springer.com/article/10.1007/s10555-014-9520-6/fulltext.html#CR10 [Accessed August 4, 2015].

Bultmann, S. et al., 2012. Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers. *Nucleic Acids Research*, 40(12), pp.5368–5377. Available at: http://nar.oxfordjournals.org/content/40/12/5368.abstract.

Byrne, S.M. et al., 2014. Multi-kilobase homozygous targeted gene replacement in human induced pluripotent stem cells. *Nucleic acids research*, p.gku1246–. Available at: http://nar.oxfordjournals.org/content/early/2014/11/20/nar.gku1246.full [Accessed February 17, 2015].

Calado, A. et al., 2002. Exportin-5-mediated nuclear export of eukaryotic elongation factor 1A and tRNA. *The EMBO Journal*, 21(22), pp.6216–6224. Available at: http://emboj.embopress.org/content/21/22/6216.abstract [Accessed August 3, 2015].

Canver, M.C. et al., 2014. Characterization of genomic deletion efficiency mediated by clustered regularly interspaced palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. *The Journal of biological chemistry*, 289(31), pp.21312–24. Available at: http://www.jbc.org/content/289/31/21312.full [Accessed July 27, 2015].

Carlson, D.F. et al., 2012. Efficient TALEN-mediated gene knockout in livestock. *Proceedings of the National Academy of Sciences*. Available at: http://www.pnas.org/content/early/2012/09/26/1211446109.abstract.

Carroll, D., 2014. Genome engineering with targetable nucleases. *Annual review of biochemistry*, 83, pp.409–39. Available at: http://www.annualreviews.org/doi/abs/10.1146/annurev-biochem-060713-035418 [Accessed May 26, 2015].

Carroll, D., 2011. Genome engineering with zinc-finger nucleases. *Genetics*, 188(4), pp.773–82. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3176093&tool=pmcentrez&rendertype=abstract [Accessed February 9, 2015].

Cartier, N. et al., 2009. Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked adrenoleukodystrophy. *Science (New York, N.Y.)*, 326(5954), pp.818–23. Available at: http://www.sciencemag.org/content/326/5954/818.abstract?sid=223c0893-daca-4cc8-a093-341572f4e2b4 [Accessed July 19, 2015].

Cerbini, T. et al., 2015. Transfection, Selection, and Colony-picking of Human Induced Pluripotent Stem Cells TALEN-targeted with a GFP Gene into the AAVS1 Safe Harbor. *Journal of visualized experiments : JoVE*, (96), p.e52504. Available at: http://www.jove.com/video/52504/transfection-selection-colony-picking-human-induced-pluripotent-stem [Accessed March 9, 2015].

Cermak, T. et al., 2011. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Research*, 39(12), p.e82. Available at: http://nar.oxfordjournals.org/content/39/12/e82.abstract.

Chalberg, T.W. et al., 2006. Integration Specificity of Phage ϕC31 Integrase in the Human Genome. *Journal of Molecular Biology*, 357(1), pp.28–48. Available at: http://www.sciencedirect.com/science/article/pii/S002228360501586X.

Chandler, K.J. et al., 2007. Relevance of BAC transgene copy number in mice: transgene copy number variation across multiple transgenic lines and correlations with transgene integrity and expression. *Mammalian genome : official journal of the International Mammalian Genome Society*, 18(10), pp.693–708. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3110064&tool=pmcentrez&rendertype=abstract [Accessed January 18, 2016].

Chang, C.-J. & Bouhassira, E.E., 2012. Zinc-finger nuclease-mediated correction of α-thalassemia in iPS cells. *Blood*, 120(19), pp.3906–3914. Available at: http://bloodjournal.hematologylibrary.org/content/120/19/3906.abstract.

Chang, R. & Wang, E., 2007. Mouse translation elongation factor eEF1A-2 interacts with Prdx-I to protect cells against apoptotic death induced by oxidative stress. *Journal of cellular biochemistry*, 100(2), pp.267–78. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16888816 [Accessed August 2, 2015].

Chen et al., 2011. A comparison of exogenous promoter activity at the ROSA26 locus using a ΦiC31 integrase mediated cassette exchange approach in mouse ES cells. *PloS one*, 6(8), p.e23376. Available at: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0023376

[Accessed February 6, 2015].

Chen, Oikonomou, G., et al., 2013. A large-scale in vivo analysis reveals that TALENs are significantly more mutagenic than ZFNs generated using context-dependent assembly. *Nucleic Acids Res*, 41(4), pp.2769–2778.

Chen et al., 2015. Engineering Human Stem Cell Lines with Inducible Gene Knockout using CRISPR/Cas9. *Cell Stem Cell*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26145478 [Accessed July 5, 2015].

Chen et al., 2000. Rapid up-regulation of peptide elongation factor EF-1alpha protein levels is an immediate early event during oxidative stress-induced apoptosis. *Experimental cell research*, 259(1), pp.140–8. Available at: http://www.sciencedirect.com/science/article/pii/S0014482700949528 [Accessed August 2, 2015].

Chen, Fenk, L.A. & de Bono, M., 2013. Efficient genome editing in Caenorhabditis elegans by CRISPR-targeted homologous recombination. *Nucleic acids research*, 41(20), p.e193. Available at: http://nar.oxfordjournals.org/content/41/20/e193.full [Accessed February 8, 2016].

Cheung, A.K. et al., 1980. Integration of the adeno-associated virus genome into cellular DNA in latently infected human Detroit 6 cells. *J. Virol.*, 33(2), pp.739–748. Available at: http://jvi.asm.org/content/33/2/739.short [Accessed June 23, 2015].

Cho, S.J. et al., 2012. Translation elongation factor-1A1 (eEF1A1) localizes to the spine by domain III. *Biochemistry and Molecular Biology Reports*, 45(4), pp.227–232.

Cho, S.W. et al., 2014. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome research*, 24(1), pp.132–41. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3875854&tool=pmcentrez&rendertype=abstract [Accessed July 11, 2014].

Choi, P.S. & Meyerson, M., 2014. Targeted genomic rearrangements using CRISPR/Cas technology. *Nature communications*, 5, p.3728. Available at: http://www.nature.com/ncomms/2014/140424/ncomms4728/full/ncomms4728.html [Accessed July 14, 2014].

Chong, C.R. et al., 2006. Identification of type 1 inosine monophosphate dehydrogenase as an antiangiogenic drug target. *Journal of medicinal chemistry*, 49(9), pp.2677–80. Available at: http://dx.doi.org/10.1021/jm051225t [Accessed May 26, 2015].

Choulika, A. et al., 1995. Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of Saccharomyces cerevisiae. *Mol. Cell. Biol.*, 15(4), pp.1968–1973. Available at: http://mcb.asm.org/content/15/4/1968.abstract?ijkey=ef533b6c35ce4d4088cd5a4a67c3bb4b618329af&keytype2=tf_ipsecsha [Accessed June 8, 2015].

Christian, M.L. et al., 2012. Targeting G with TAL Effectors: A Comparison of Activities of TALENs Constructed with NN and NK Repeat Variable Di-Residues. *PLoS ONE*, 7(9), p.e45383. Available at: http://dx.doi.org/10.1371/journal.pone.0045383.

Chu, H.W. et al., 2015. CRISPR-Cas9 mediated gene knockout in primary human airway epithelial cells reveals a pro-inflammatory role for MUC18. *Gene therapy*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26043872 [Accessed June 9, 2015].

Church, G., 2013. George Church: De-Extinction Is a Good Idea. *Scientific American*, 309(3), pp.12–12. Available at: http://www.scientificamerican.com/article/george-church-de-extinction-is-a-good-idea/ [Accessed June 5, 2015].

Condeelis, J., 1995. Elongation factor 1α, translation and the cytoskeleton. *Trends in Biochemical Sciences*, 20(5), pp.169–170. Available at: http://www.sciencedirect.com/science/article/pii/S0968000400889987 [Accessed July 24, 2015].

Cong, L. et al., 2012. Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nature communications*, 3, p.968. Available at: http://dx.doi.org/10.1038/ncomms1962 [Accessed July 24, 2015].

Cong, L. et al., 2013. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science*, 339(6121), pp.819–823. Available at: http://www.sciencemag.org/content/339/6121/819.abstract.

Decottignies, A., 2014. [The telomere position effect: silence in the back row!]. *Médecine sciences : M/S*, 30(2), pp.173–8. Available at: http://www.medecinesciences.org/articles/medsci/full_html/2014/02/medsci20143002p173/medsci20143002p173.html [Accessed June 17, 2015].

DeKelver, R.C. et al., 2010. Functional genomics, proteomics, and regulatory DNA analysis in isogenic settings using zinc finger nuclease-driven transgenesis into a safe harbor locus in the human genome. *Genome Research*, 20(8), pp.1133–1142. Available at: http://genome.cshlp.org/content/20/8/1133.abstract.

Dhalluin, C. et al., 1999. Structure and ligand of a histone acetyltransferase bromodomain. *Nature*, 399(6735), pp.491–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10365964 [Accessed June 19, 2015].

Digits, J.A. & Hedstrom, L., 1999. Species-Specific Inhibition of Inosine 5‘-Monophosphate Dehydrogenase by Mycophenolic Acid†. *Biochemistry*, 38(46), pp.15388–15397. Available at: http://dx.doi.org/10.1021/bi991558q.

Dillon, N. & Festenstein, R., 2002. Unravelling heterochromatin: competition between positive and negative factors regulates accessibility. *Trends in Genetics*, 18(5), pp.252–258. Available at: http://www.sciencedirect.com/science/article/pii/S0168952502026483 [Accessed June 19, 2015].

Ding, Q. et al., 2013. Enhanced efficiency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs. *Cell stem cell*, 12(4), pp.393–4. Available at:

http://www.sciencedirect.com/science/article/pii/S193459091300101X [Accessed August 27, 2014].

Dutheil, N. et al., 2004. Characterization of the Mouse Adeno-Associated Virus AAVS1 Ortholog. *Journal of Virology*, 78(16), pp.8917–8921. Available at: http://jvi.asm.org/content/78/16/8917.abstract.

Ebina, H. et al., 2015. A high excision potential of TALENs for integrated DNA of HIV-based lentiviral vector. *PloS one*, 10(3), p.e0120047. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4363575&tool=pmcentrez&rendertype=abstract [Accessed June 18, 2015].

Elliott, B. et al., 1998. Gene Conversion Tracts from Double-Strand Break Repair in Mammalian Cells. *Mol. Cell. Biol.*, 18(1), pp.93–101. Available at: http://mcb.asm.org/content/18/1/93.long [Accessed June 8, 2015].

Emery, D.W. et al., 2000. A chromatin insulator protects retrovirus vectors from chromosomal position effects. *Proceedings of the National Academy of Sciences of the United States of America*, 97(16), pp.9150–5. Available at: http://www.pnas.org/content/97/16/9150.abstract [Accessed January 18, 2016].

Engler, C. et al., 2009. Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes. *PLoS ONE*, 4(5), p.e5553. Available at: http://dx.doi.org/10.1371/journal.pone.0005553.

Engler, C., Kandzia, R. & Marillonnet, S., 2008. A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLoS ONE*, 3(11), p.e3647. Available at: http://dx.doi.org/10.1371/journal.pone.0003647.

Esvelt, K.M. et al., 2013. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nature methods*, 10(11), pp.1116–21. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3844869&tool=pmcentrez&rendertype=abstract [Accessed July 10, 2014].

Farazi, T. et al., 1997. Isolation and Characterization of Mycophenolic Acid-resistant Mutants of Inosine-5′-monophosphate Dehydrogenase. *Journal of Biological Chemistry*, 272(2), pp.961–965. Available at:

http://www.jbc.org/content/272/2/961.abstract.

Farson, D. et al., 2001. A new-generation stable inducible packaging cell line for lentiviral vectors. *Human gene therapy*, 12(8), pp.981–97. Available at: http://online.liebertpub.com/doi/abs/10.1089/104303401750195935 [Accessed July 24, 2015].

Feng, Y.Q. et al., 2001. Position effects are influenced by the orientation of a transgene with respect to flanking chromatin. *Molecular and cellular biology*, 21(1), pp.298–309. Available at: http://mcb.asm.org/content/21/1/298.short [Accessed June 21, 2015].

Fleming, M.A. et al., 1996. Inhibition of IMPDH by Mycophenolic Acid: Dissection of Forward and Reverse Pathways Using Capillary Electrophoresis. *Biochemistry*, 35(22), pp.6990–6997. Available at: http://dx.doi.org/10.1021/bi9607416.

Flemr, M. & Bühler, M., 2015. Single-Step Generation of Conditional Knockout Mouse Embryonic Stem Cells. *Cell reports*, 12(4), pp.709–16. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26190102 [Accessed July 22, 2015].

Fontes, A. & Lakshmipathy, U., 2013. Advances in genetic modification of pluripotent stem cells. *Biotechnology Advances*, 31(7), pp.994–1001. Available at: http://www.sciencedirect.com/science/article/pii/S0734975013001171.

Friedrich, G. & Soriano, P., 1991. Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes & development*, 5(9), pp.1513–23. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1653172 [Accessed July 24, 2015].

Fu, Y. et al., 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol*.

Fu, Y. et al., 2014. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nature biotechnology*, 32(3), pp.279–84. Available at: http://dx.doi.org/10.1038/nbt.2808 [Accessed July 9, 2014].

Fung, H. & Weinstock, D.M., 2011. Repair at Single Targeted DNA Double-Strand

Breaks in Pluripotent and Differentiated Human Cells. *PLoS ONE*, 6(5), p.e20514. Available at: http://dx.doi.org/10.1371/journal.pone.0020514.

Gaj, T. et al., 2012. Targeted gene knockout by direct delivery of zinc-finger nuclease proteins. *Nat Methods*, 9(8), pp.805–807.

Gaj, T., Gersbach, C.A. & Barbas Iii, C.F., 2013. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in Biotechnology*, (0). Available at: http://www.sciencedirect.com/science/article/pii/S0167779913000875.

Gama-Norton, L. et al., 2011. Lentivirus production is influenced by SV40 large T-antigen and chromosomal integration of the vector in HEK293 cells. *Human gene therapy*, 22(10), pp.1269–79. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21554103 [Accessed May 27, 2015].

Garrick, D. et al., 1998. Repeat-induced gene silencing in mammals. *Nature genetics*, 18(1), pp.56–9. Available at: http://dx.doi.org/10.1038/ng0198-56 [Accessed May 28, 2015].

Gilbert, L.A. et al., 2013. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*, 154(2), pp.442–51. Available at: http://www.sciencedirect.com/science/article/pii/S009286741300826X [Accessed July 11, 2014].

Giraldo, P. & Montoliu, L., 2001. Size matters: use of YACs, BACs and PACs in transgenic animals. *Transgenic research*, 10(2), pp.83–103. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11305364 [Accessed January 26, 2016].

Gonen, H. et al., 1994. Protein synthesis elongation factor EF-1 alpha is essential for ubiquitin-dependent degradation of certain N alpha-acetylated proteins and may be substituted for by the bacterial elongation factor EF-Tu. *Proceedings of the National Academy of Sciences*, 91(16), pp.7648–7652. Available at: http://www.pnas.org/content/91/16/7648.abstract [Accessed August 3, 2015].

Gong, S. et al., 2003. A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. *Nature*, 425(6961), pp.917–925. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14586460 [Accessed October 8, 2015].

Gordon, J.W. et al., 1980. Genetic transformation of mouse embryos by microinjection of purified DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 77(12), pp.7380–4. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=350507&tool=pmcentrez&rendertype=abstract [Accessed May 17, 2015].

Gratz, S.J. et al., 2014. Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila. *Genetics*, 196(4), pp.961–71. Available at: http://www.genetics.org/content/196/4/961.full [Accessed July 10, 2014].

Gray, S.J. et al., 2011. Optimizing promoters for recombinant adeno-associated virus-mediated gene expression in the peripheral and central nervous system using self-complementary vectors. *Human gene therapy*, 22(9), pp.1143–53. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3177952&tool=pmcentrez&rendertype=abstract [Accessed June 25, 2015].

Greene, M.R. et al., 2012. Transduction of human CD34+ repopulating cells with a self-inactivating lentiviral vector for SCID-X1 produced at clinical scale by a stable cell line. *Human gene therapy methods*, 23(5), pp.297–308. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3732136&tool=pmcentrez&rendertype=abstract [Accessed July 24, 2015].

Gu, J.J. et al., 2003. Targeted disruption of the inosine 5'-monophosphate dehydrogenase type I gene in mice. *Mol Cell Biol*, 23(18), pp.6702–6712.

Guilinger, J.P., Pattanayak, V., et al., 2014. Broad specificity profiling of TALENs results in engineered nucleases with improved DNA-cleavage specificity. *Nature methods*, 11(4), pp.429–35. Available at: http://www.nature.com.ezproxy.is.ed.ac.uk/nmeth/journal/v11/n4/full/nmeth.2845.html [Accessed June 23, 2015].

Guilinger, J.P., Thompson, D.B. & Liu, D.R., 2014. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat Biotechnol*.

Gupta, A. et al., 2013. Targeted chromosomal deletions and inversions in zebrafish. *Genome research*, 23(6), pp.1008–17. Available at: http://genome.cshlp.org/content/23/6/1008.abstract?ijkey=621c501cc7ef9a2416ec8e263aee4ac82c63e773&keytype2=tf_ipsecsha [Accessed July 24, 2015].

Hacein-Bey-Abina, S. et al., 2008. Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *The Journal of clinical investigation*, 118(9), pp.3132–42. Available at: http://www.jci.org/articles/view/35700 [Accessed March 29, 2015].

Han, J. et al., 2014. Efficient in vivo deletion of a large imprinted lncRNA by CRISPR/Cas9. *RNA biology*, 11(7), pp.829–35. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4179957&tool=pmcentrez&rendertype=abstract [Accessed July 27, 2015].

Harrison, M.M. et al., 2014. A CRISPR view of development. *Genes & Development*, 28(17), pp.1859–1872. Available at: http://genesdev.cshlp.org/content/28/17/1859.full [Accessed September 2, 2014].

Hashimoto, K. & Ishima, T., 2011. Neurite outgrowth mediated by translation elongation factor eEF1A1: a target for antiplatelet agent cilostazol. *PloS one*, 6(3), p.e17431. Available at: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0017431 [Accessed July 24, 2015].

Hastie, E. & Samulski, R.J., 2015. Adeno-associated virus at 50: a golden anniversary of discovery, research, and gene therapy success-a personal perspective. *Human gene therapy*, 26(5), pp.257–65. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25807962 [Accessed May 22, 2015].

He, J., Yang, Q. & Chang, L.-J., 2005. Dynamic DNA methylation and histone modifications contribute to lentiviral transgene silencing in murine embryonic carcinoma cells. *Journal of virology*, 79(21), pp.13497–508. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1262567&tool=pmcentrez&rendertype=abstract [Accessed September 2, 2015].

Hedstrom, L., 2009. IMP Dehydrogenase: Structure, Mechanism, and Inhibition. *Chemical Reviews*, 109(7), pp.2903–2928. Available at: http://dx.doi.org/10.1021/cr900021w.

Henckaerts, E. et al., 2009. Site-specific integration of adeno-associated virus involves partial duplication of the target locus. *Proceedings of the National Academy of Sciences of the United States of America*, 106(18), pp.7571–6. Available at: http://www.pnas.org/content/106/18/7571.full [Accessed July 26, 2015].

Hirata, R. et al., 2002. Targeted transgene insertion into human chromosomes by adeno-associated virus vectors. *Nature biotechnology*, 20(7), pp.735–8. Available at: http://dx.doi.org/10.1038/nbt0702-735 [Accessed June 16, 2015].

Hockemeyer, D. et al., 2011. Genetic engineering of human pluripotent cells using TALE nucleases. *Nature biotechnology*, 29(8), pp.731–4. Available at: http://dx.doi.org/10.1038/nbt.1927 [Accessed July 11, 2014].

Hodges, S.D. et al., 1989. Increased activity, amount, and altered kinetic properties of IMP dehydrogenase from mycophenolic acid-resistant neuroblastoma cells. *Journal of Biological Chemistry*, 264(30), pp.18137–18141. Available at: http://www.jbc.org/content/264/30/18137.abstract.

Hoggan, M.D., Blacklow, N.R. & Rowe, W.P., 1966. Studies of small DNA viruses found in various adenovirus preparations: physical, biological, and immunological characteristics. *Proceedings of the National Academy of Sciences of the United States of America*, 55(6), pp.1467–74. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=224346&tool=pmcentrez&rendertype=abstract [Accessed June 23, 2015].

Hotokezaka, Y. et al., 2002. Interaction of the eukaryotic elongation factor 1A with newly synthesized polypeptides. *The Journal of biological chemistry*, 277(21), pp.18545–51. Available at: http://www.jbc.org/content/277/21/18545.abstract [Accessed August 3, 2015].

Hou, Z. et al., 2013. Efficient genome engineering in human pluripotent stem cells

using Cas9 from Neisseria meningitidis. *Proceedings of the National Academy of Sciences of the United States of America*, 110(39), pp.15644–9. Available at: http://www.pnas.org/content/110/39/15644.abstract [Accessed August 4, 2015].

Houdebine, L.-M., 2002. The methods to generate transgenic animals and to control transgene expression. *Journal of Biotechnology*, 98(2-3), pp.145–160. Available at: http://www.sciencedirect.com/science/article/pii/S0168165602001293 [Accessed June 15, 2015].

Hruscha, A. et al., 2013. Efficient CRISPR/Cas9 genome editing with low off-target effects in zebrafish. *Development (Cambridge, England)*, 140(24), pp.4982–7. Available at: http://dev.biologists.org/content/140/24/4982.abstract?sid=27503590-dc41-4e52-9510-da0ecd90c5c4 [Accessed July 23, 2014].

Hsu, P.D. et al., 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol*, 31(9), pp.827–832.

Hu, W. et al., 2014. RNA-directed gene editing specifically eradicates latent and prevents new HIV-1 infection. *Proc Natl Acad Sci U S A*.

Iiizumi, S. et al., 2008. Impact of non-homologous end-joining deficiency on random and targeted DNA integration: implications for gene targeting. *Nucleic acids research*, 36(19), pp.6333–42. Available at: http://nar.oxfordjournals.org/content/36/19/6333.abstract?ijkey=bfa1f6145f8ad5a0d700aaf43fdef8c30b4fef91&keytype2=tf_ipsecsha [Accessed June 8, 2015].

Ioshikhes, I.P. & Zhang, M.Q., 2000. Large-scale human promoter mapping using CpG islands. *Nature genetics*, 26(1), pp.61–3. Available at: http://dx.doi.org/10.1038/79189 [Accessed June 20, 2015].

Irion, S. et al., 2007. Identification and targeting of the ROSA26 locus in human embryonic stem cells. *Nature Biotechnology*, 25(12), pp.1477–1482. Available at: http://dx.doi.org/10.1038/nbt1362.

Ishikawa, Y. et al., 2006. Phage phiC31 integrase-mediated genomic integration of the common cytokine receptor gamma chain in human T-cell lines. *The journal*

of gene medicine, 8(5), pp.646–53. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16508910 [Accessed January 19, 2016].

Ivics, Z. et al., 1997. Molecular Reconstruction of Sleeping Beauty, a Tc1-like Transposon from Fish, and Its Transposition in Human Cells. *Cell*, 91(4), pp.501–510. Available at: http://www.sciencedirect.com/science/article/pii/S0092867400804365.

Jain, J. et al., 2001. VX-497: A novel, selective IMPDH inhibitor and immunosuppressive agent. *Journal of Pharmaceutical Sciences*, 90(5), pp.625–637. Available at: http://doi.wiley.com/10.1002/1520-6017%28200105%2990%3A5%3C625%3A%3AAID-JPS1019%3E3.0.CO%3B2-1 [Accessed July 17, 2015].

Jinek, M. et al., 2012. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, 337(6096), pp.816–821. Available at: http://www.sciencemag.org/content/337/6096/816.abstract.

Jones, P.L. et al., 1998. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nature genetics*, 19(2), pp.187–91. Available at: http://dx.doi.org/10.1038/561 [Accessed June 20, 2015].

Kahns, S., 1998. The elongation factor 1 A-2 isoform from rabbit: cloning of the cDNA and characterization of the protein. *Nucleic Acids Research*, 26(8), pp.1884–1890. Available at: http://nar.oxfordjournals.org/content/26/8/1884.full [Accessed July 30, 2015].

Kamakaka, R.T. & Biggins, S., 2005. Histone variants: deviants? *Genes & development*, 19(3), pp.295–310. Available at: http://genesdev.cshlp.org/content/19/3/295.long [Accessed August 16, 2015].

Kamionka, M., 2011. Engineering of therapeutic proteins production in Escherichia coli. *Current pharmaceutical biotechnology*, 12(2), pp.268–74. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3179032&tool=pmcentrez&rendertype=abstract [Accessed July 10, 2015].

Kawamura, M. et al., 2014. The Prognostic Significance of Eukaryotic Elongation

Factor 1 Alpha-2 in Non-Small Cell Lung Cancer. *Anticancer Res*, 34(2), pp.651–658. Available at: http://ar.iiarjournals.org/content/34/2/651.abstract [Accessed July 29, 2015].

Khacho, M. et al., 2008. eEF1A is a novel component of the mammalian nuclear protein export machinery. *Molecular biology of the cell*, 19(12), pp.5296–308. Available at: http://www.molbiolcell.org/content/19/12/5296.full [Accessed July 19, 2015].

Khan, I.F., Hirata, R.K. & Russell, D.W., 2011. AAV-mediated gene targeting methods for human cells. *Nature protocols*, 6(4), pp.482–501. Available at: http://www.nature.com.ezproxy.is.ed.ac.uk/nprot/journal/v6/n4/full/nprot.2011.301.html [Accessed June 23, 2015].

Kim, J. et al., 2009. The role of translation elongation factor eEF1A in intracellular alkalinization-induced tumor cell growth. *Laboratory investigation; a journal of technical methods and pathology*, 89(8), pp.867–74. Available at: http://dx.doi.org/10.1038/labinvest.2009.53 [Accessed July 28, 2015].

Kim, Y.G., Cha, J. & Chandrasegaran, S., 1996. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences*, 93(3), pp.1156–1160. Available at: http://www.pnas.org/content/93/3/1156.abstract [Accessed August 13, 2015].

Kleinstiver, B.P. et al., 2015. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature*, advance on. Available at: http://dx.doi.org/10.1038/nature14592 [Accessed June 22, 2015].

Knight, S. et al., 2012. Safer, silencing-resistant lentiviral vectors: optimization of the ubiquitous chromatin-opening element through elimination of aberrant splicing. *Journal of virology*, 86(17), pp.9088–95. Available at: http://jvi.asm.org/content/86/17/9088.full [Accessed December 8, 2015].

Knudsen, S.M. et al., 1993. Tissue-dependent variation in the expression of elongation factor-1α isoforms: Isolation and characterisation of a cDNA encoding a novel variant of human elongation-factor 1α. *European Journal of*

Biochemistry, 215(3), pp.549–554. Available at: http://dx.doi.org/10.1111/j.1432-1033.1993.tb18064.x.

Kobayashi, T. et al., 2012. Identification of rat Rosa26 locus enables generation of knock-in rat lines ubiquitously expressing tdTomato. *Stem cells and development*, 21(16), pp.2981–6. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3475147&tool=pmcentrez&rendertype=abstract [Accessed September 11, 2014].

Koike-Yusa, H. et al., 2013. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat Biotechnol*.

Kotin, R.M., Linden, R.M. & Berns, K.I., 1992. Characterization of a preferred site on human chromosome 19q for integration of adeno-associated virus DNA by non-homologous recombination. *The EMBO journal*, 11(13), pp.5071–5078.

Krastel, P. et al., 2015. Nannocystin A: an Elongation Factor 1 Inhibitor from Myxobacteria with Differential Anti-Cancer Properties. *Angewandte Chemie*, p.n/a–n/a. Available at: http://doi.wiley.com/10.1002/ange.201505069 [Accessed August 4, 2015].

Krishnan, M. et al., 2006. Effects of epigenetic modulation on reporter gene expression: implications for stem cell imaging. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 20(1), pp.106–8. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3625424&tool=pmcentrez&rendertype=abstract [Accessed June 16, 2015].

Lachner, M. et al., 2001. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature*, 410(6824), pp.116–20. Available at: http://dx.doi.org/10.1038/35065132 [Accessed March 3, 2015].

Lee, C.M. et al., 2012. Correction of the DeltaF508 Mutation in the Cystic Fibrosis Transmembrane Conductance Regulator Gene by Zinc-Finger Nuclease Homology-Directed Repair. *Biores Open Access*, 1(3), pp.99–108.

Lee, H.J., Kim, E. & Kim, J.-S., 2010. Targeted chromosomal deletions in human

cells using zinc finger nucleases. *Genome Research*, 20(1), pp.81–89. Available at: http://genome.cshlp.org/content/20/1/81.abstract.

Lee, S., Wolfraim, L.A. & Wang, E., 1993. Differential expression of S1 and elongation factor-1 alpha during rat development. *The Journal of biological chemistry*, 268(32), pp.24453–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8226996 [Accessed August 1, 2015].

Li, D. et al., 2013. The Unexpected Roles of Eukaryotic Translation Elongation Factors in RNA Virus Replication and Pathogenesis. *Microbiology and Molecular Biology Reviews*, 77(2), pp.253–266. Available at: http://mmbr.asm.org/content/77/2/253.abstract.

Li, L., Wu, L.P. & Chandrasegaran, S., 1992. Functional domains in Fok I restriction endonuclease. *Proceedings of the National Academy of Sciences*, 89(10), pp.4275–4279. Available at: http://www.pnas.org/content/89/10/4275.abstract [Accessed August 13, 2015].

Li, X. et al., 2001. Influence of DNA methylation on transgene expression. *Chinese Science Bulletin*, 46(15), pp.1300–1303. Available at: http://link.springer.com/10.1007/BF03184330 [Accessed June 20, 2015].

Li, X. et al., 2014. Rosa26-targeted swine models for stable gene over-expression and Cre-mediated lineage tracing. *Cell research*, 24(4), pp.501–4. Available at: http://dx.doi.org/10.1038/cr.2014.15 [Accessed October 14, 2014].

Li, Y. et al., 2015. Excision of Expanded GAA Repeats Alleviates the Molecular Phenotype of Friedreich's Ataxia. *Molecular therapy : the journal of the American Society of Gene Therapy*, 23(6), pp.1055–65. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25758173 [Accessed August 15, 2015].

Li, Z. et al., 2010. Eef1a2 promotes cell growth, inhibits apoptosis and activates JAK/STAT and AKT signaling in mouse plasmacytomas. *PloS one*, 5(5), p.e10755. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2873962&tool=pmcentrez&rendertype=abstract [Accessed August 2, 2015].

Liang, F. et al., 1996. Chromosomal double-strand break repair in Ku80-deficient cells. *Proceedings of the National Academy of Sciences*, 93(17), pp.8929–8933. Available at: http://www.pnas.org/content/93/17/8929.abstract?ijkey=8b31abd53cc4d67efb58d8e2e71c9288773166cc&keytype2=tf_ipsecsha [Accessed June 8, 2015].

Liang, P. et al., 2015. CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes. *Protein & cell*, 6(5), pp.363–72. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4417674&tool=pmcentrez&rendertype=abstract [Accessed April 23, 2015].

de Ligt, J. et al., 2012. Diagnostic Exome Sequencing in Persons with Severe Intellectual Disability. *New England Journal of Medicine*, 367(20), pp.1921–1929. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23033978 [Accessed July 6, 2015].

Lin, W.-S. et al., 2012. Hepatitis B virus X protein blocks filamentous actin bundles by interaction with eukaryotic translation elongat ion factor 1 alpha 1. *Journal of medical virology*, 84(6), pp.871–7. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22499008 [Accessed August 2, 2015].

Link, J.O. & Straub, K., 1996. Trapping of an IMP Dehydrogenase−Substrate Covalent Intermediate by Mycophenolic Acid. *Journal of the American Chemical Society*, 118(8), pp.2091–2092. Available at: http://dx.doi.org/10.1021/ja9534056.

Liu, G. et al., 2005. Target-site Preferences of Sleeping Beauty Transposons. *Journal of Molecular Biology*, 346(1), pp.161–173. Available at: http://www.sciencedirect.com/science/article/pii/S0022283604012574.

Liu, J. et al., 2009. Stable transgene expression in human embryonic stem cells after simple chemical transfection. *Molecular reproduction and development*, 76(6), pp.580–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19034957 [Accessed January 18, 2016].

Liu, R. et al., 1996. Homozygous Defect in HIV-1 Coreceptor Accounts for

Resistance of Some Multiply-Exposed Individuals to HIV-1 Infection. *Cell*, 86(3), pp.367–377. Available at: http://www.sciencedirect.com/science/article/pii/S0092867400801105 [Accessed August 15, 2015].

Luger, K. et al., 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, 389(6648), pp.251–60. Available at: http://dx.doi.org/10.1038/38444 [Accessed July 17, 2014].

Lund, A. et al., 1996. Assignment of Human Elongation Factor 1α Genes:EEF1AMaps to Chromosome 6q14 andEEF1A2to 20q13.3. *Genomics*, 36(2), pp.359–361. Available at: http://www.sciencedirect.com/science/article/pii/S0888754396904759.

Madsen, H.O. et al., 1990. Retropseudogenes constitute the major part of the human elongation factor 1α gene family. *Nucleic Acids Research*, 18(6), pp.1513–1516. Available at: http://nar.oxfordjournals.org/content/18/6/1513.abstract.

Maggio, I. et al., 2014. Adenoviral vector delivery of RNA-guided CRISPR/Cas9 nuclease complexes induces targeted mutagenesis in a diverse array of human cells. *Sci Rep*, 4, p.5105.

Mak, A.N.-S. et al., 2012. The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science (New York, N.Y.)*, 335(6069), pp.716–9. Available at: http://www.sciencemag.org/content/335/6069/716.full?sid=6e9cfd33-973a-4dbd-8ce3-57ab7045d046 [Accessed July 24, 2015].

Mali, P., Aach, J., et al., 2013. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol*, 31(9), pp.833–838.

Mali, P., Yang, L., et al., 2013. RNA-Guided Human Genome Engineering via Cas9. *Science*, 339(6121), pp.823–826. Available at: http://www.sciencemag.org/content/339/6121/823.abstract.

Mali, P., Esvelt, K.M. & Church, G.M., 2013. Cas9 as a versatile tool for engineering biology. *Nature methods*, 10(10), pp.957–63. Available at:

http://www.nature.com/nmeth/journal/v10/n10/full/nmeth.2649.html#ref19 [Accessed July 9, 2014].

Mandal, P.K. et al., 2014. Efficient Ablation of Genes in Human Hematopoietic Stem and Effector Cells using CRISPR/Cas9. *Cell Stem Cell*, 15(5), pp.643–652. Available at: http://www.cell.com/article/S193459091400455X/fulltext [Accessed November 7, 2014].

Mateyak, M.K. & Kinzy, T.G., 2010. eEF1A: thinking outside the ribosome. *The Journal of biological chemistry*, 285(28), pp.21209–13. Available at: http://www.jbc.org/content/285/28/21209.abstract?sid=4f7389a3-0309-4a69-abc5-bb6a9da598a0 [Accessed July 24, 2015].

Matteo, M. et al., 2012. PiggyBac Toolbox. In Y. Bigot, ed. *Mobile Genetic Elements*. Humana Press, pp. 241–254. Available at: http://dx.doi.org/10.1007/978-1-61779-603-6_14.

McCready, P.M. et al., 1997. Multiple negative and positive cis-acting elements control the expression of the murine CD4 gene. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, 1351(1-2), pp.181–191. Available at: http://www.sciencedirect.com/science/article/pii/S0167478196001947 [Accessed July 15, 2015].

McInerney, J.M., Nawrocki, J.R. & Lowrey, C.H., 2000. Long-term silencing of retroviral vectors is resistant to reversal by trichostatin A and 5-azacytidine. *Gene therapy*, 7(8), pp.653–63. Available at: http://www.nature.com/gt/journal/v7/n8/full/3301155a.html [Accessed February 8, 2016].

Meckler, J.F. et al., 2013. Quantitative analysis of TALE-DNA interactions suggests polarity effects. *Nucleic acids research*, 41(7), pp.4118–28. Available at: http://nar.oxfordjournals.org/content/41/7/4118.abstract [Accessed August 7, 2015].

Menon, T. et al., 2015. Lymphoid Regeneration from Gene-Corrected SCID-X1 Subject-Derived iPSCs. *Cell Stem Cell*, 16(4), pp.367–372. Available at:

http://www.sciencedirect.com/science/article/pii/S1934590915000612 [Accessed March 13, 2015].

Merrihew, R. et al., 1996. High-frequency illegitimate integration of transfected DNA at preintegrated target sites in a mammalian genome. *Mol. Cell. Biol.*, 16(1), pp.10–18. Available at: http://mcb.asm.org/content/16/1/10.abstract?ijkey=0b73f0b373ab07573b50f64b4dfdc27c7c78f016&keytype2=tf_ipsecsha [Accessed July 28, 2015].

Miller, J.C. et al., 2007. An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat Biotechnol*, 25(7), pp.778–785.

Mitchell, R.S. et al., 2004. Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS biology*, 2(8), p.E234. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=509299&tool=pmcentrez&rendertype=abstract [Accessed May 31, 2015].

Mock, U. et al., 2015. mRNA transfection of a novel TAL effector nuclease (TALEN) facilitates efficient knockout of HIV co-receptor CCR5. *Nucleic acids research*, 43(11), pp.5560–5571. Available at: http://nar.oxfordjournals.org/content/43/11/5560 [Accessed May 25, 2015].

Moehle, E.A. et al., 2007. Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proceedings of the National Academy of Sciences*, 104(9), pp.3055–3060. Available at: http://www.pnas.org/content/104/9/3055.abstract.

Morano, A. et al., 2014. Targeted DNA methylation by homology-directed repair in mammalian cells. Transcription reshapes methylation on the repaired gene. *Nucleic acids research*, 42(2), pp.804–21. Available at: http://nar.oxfordjournals.org/content/42/2/804.long [Accessed May 4, 2015].

Moscou, M.J. & Bogdanove, A.J., 2009. A Simple Cipher Governs DNA Recognition by TAL Effectors. *Science*, 326(5959), p.1501. Available at: http://www.sciencemag.org/content/326/5959/1501.abstract.

Muller, H.J., 1930. Types of visible variations induced by X-rays inDrosophila.

Journal of Genetics, 22(3), pp.299–334. Available at: http://link.springer.com/10.1007/BF02984195 [Accessed May 6, 2015].

Müller-Kuller, U. et al., 2015. A minimal ubiquitous chromatin opening element (UCOE) effectively prevents silencing of juxtaposed heterologous promoters by epigenetic remodeling in multipotent and pluripotent stem cells. *Nucleic acids research*, 43(3), pp.1577–92. Available at: http://nar.oxfordjournals.org/content/early/2015/01/20/nar.gkv019 [Accessed December 8, 2015].

Muramatsu, D. et al., 2013. Pericentric heterochromatin generated by HP1 protein interaction-defective histone methyltransferase Suv39h1. *The Journal of biological chemistry*, 288(35), pp.25285–96. Available at: http://www.jbc.org/content/288/35/25285.long [Accessed May 29, 2015].

Murray, J.W. et al., 1996. Bundling of actin filaments by elongation factor 1 alpha inhibits polymerization at filament ends. *The Journal of cell biology*, 135(5), pp.1309–21. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2121097&tool=pmcentrez&rendertype=abstract [Accessed July 7, 2015].

Mussolino, C. et al., 2011. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Research*, 39(21), pp.9283–9293. Available at: http://nar.oxfordjournals.org/content/39/21/9283.abstract.

Mussolino, C. et al., 2014. TALENs facilitate targeted genome editing in human cells with high specificity and low cytotoxicity. *Nucleic acids research*, 42(10), pp.6762–73. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4041469&tool=pmcentrez&rendertype=abstract [Accessed June 23, 2015].

Mutskov, V. & Felsenfeld, G., 2004. Silencing of transgene transcription precedes methylation of promoter DNA and histone H3 lysine 9. *The EMBO journal*, 23(1), pp.138–49. Available at: http://emboj.embopress.org/content/23/1/138.abstract [Accessed May 27, 2015].

Nair, V. & Shu, Q., 2007. Inosine Monophosphate Dehydrogenase as a Probe in Antiviral Drug Discovery. *Antiviral Chemistry and Chemotherapy*, 18(5), pp.245–258. Available at: http://avc.sagepub.com/content/18/5/245.abstract [Accessed July 20, 2015].

Naitou, A. et al., 2015. Heterodimeric TALENs induce targeted heritable mutations in the crustacean Daphnia magna. *Biology open*, 4(3), pp.364–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4359742&tool=pmcentrez&rendertype=abstract [Accessed July 22, 2015].

Nakajima, J. et al., 2015. De novo EEF1A2 mutations in patients with characteristic facial features, intellectual disability, autistic behaviors and epilepsy. *Clinical genetics*, 87(4), pp.356–61. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24697219 [Accessed August 2, 2015].

Nan, X. et al., 1998. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature*, 393(6683), pp.386–9. Available at: http://dx.doi.org/10.1038/30764 [Accessed June 20, 2015].

Natsumeda, Y. et al., 1990. Two distinct cDNAs for human IMP dehydrogenase. *J. Biol. Chem.*, 265(9), pp.5292–5295. Available at: http://www.jbc.org/content/265/9/5292.abstract [Accessed June 8, 2015].

Newbery, H.J. et al., 2007. Translation elongation factor eEF1A2 is essential for post-weaning survival in mice. *The Journal of biological chemistry*, 282(39), pp.28951–9. Available at: http://www.jbc.org/content/282/39/28951.short [Accessed August 1, 2015].

Neyts, J., Andrei, G. & De Clercq, E., 1998. The Novel Immunosuppressive Agent Mycophenolate Mofetil Markedly Potentiates the Antiherpesvirus Activities of Acyclovir, Ganciclovir, and Penciclovir In Vitro and In Vivo. *Antimicrob. Agents Chemother.*, 42(2), pp.216–222. Available at: http://aac.asm.org/content/42/2/216.full [Accessed July 17, 2015].

Ni, Y. et al., 2005. Generation of a packaging cell line for prolonged large-scale production of high-titer HIV-1-based lentiviral vector. *The journal of gene*

medicine, 7(6), pp.818–34. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15693055 [Accessed July 24, 2015].

Nisole, S. & Saïb, A., 2004. Early steps of retrovirus replicative cycle. *Retrovirology*, 1(1), p.9. Available at: http://www.retrovirology.com/content/1/1/9 [Accessed May 14, 2015].

Ogata, T., Kozuka, T. & Kanda, T., 2003. Identification of an insulator in AAVS1, a preferred region for integration of adeno-associated virus DNA. *Journal of Virology*, 77(16), pp.9000–9007.

Oláh, E. et al., 2006. Modulation of cancer pathways by inhibitors of guanylate metabolism. *Advances in Enzyme Regulation*, 46(1), pp.176–190. Available at: http://www.sciencedirect.com/science/article/pii/S0065257106000045.

Osborn, M.J. et al., 2013. TALEN-based gene correction for epidermolysis bullosa. *Molecular therapy : the journal of the American Society of Gene Therapy*, 21(6), pp.1151–9. Available at: http://dx.doi.org/10.1038/mt.2013.56 [Accessed June 17, 2015].

Pabo, C.O., Peisach, E. & Grant, R.A., 2001. Design and Selection of Novel Cys2His2 Zinc Finger Proteins. *Annual Review of Biochemistry*, 70(1), pp.313–340. Available at: http://www.annualreviews.org/doi/abs/10.1146/annurev.biochem.70.1.313.

Pannell, D. & Ellis, J., 2001. Silencing of gene expression: implications for design of retrovirus vectors. *Reviews in medical virology*, 11(4), pp.205–17. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11479927 [Accessed June 21, 2015].

Papadakis, E.D. et al., 2004. Promoters and control elements: designing expression cassettes for gene therapy. *Current gene therapy*, 4(1), pp.89–113. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15032617 [Accessed August 7, 2015].

Pattanayak, V. et al., 2013. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol*, 31(9), pp.839–843.

Pecorari, L. et al., 2009. Elongation Factor 1 alpha interacts with phospho-Akt in

breast cancer cells and regulates their proliferation, survival and motility. *Molecular cancer*, 8, p.58. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2727493&tool=pmcentrez&rendertype=abstract [Accessed August 2, 2015].

Perez, E.E. et al., 2008. Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nature Biotechnology*, 26(7), pp.808–816.

Perez-Pinera, P. et al., 2013. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nature methods*, 10(10), pp.973–6. Available at: http://dx.doi.org/10.1038/nmeth.2600 [Accessed August 22, 2014].

Perez-Pinera, P., Ousterout, D.G. & Gersbach, C.A., 2012. Advances in targeted genome editing. *Current Opinion in Chemical Biology*, 16(3–4), pp.268–277. Available at: http://www.sciencedirect.com/science/article/pii/S1367593112000762.

Pfaff, N. et al., 2013. A ubiquitous chromatin opening element prevents transgene silencing in pluripotent stem cells and their differentiated progeny. *Stem cells (Dayton, Ohio)*, 31(3), pp.488–99. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23307570 [Accessed December 8, 2015].

Pikaart, M.J., Recillas-Targa, F. & Felsenfeld, G., 1998. Loss of transcriptional activity of a transgene is accompanied by DNA methylation and histone deacetylation and is prevented by insulators. *Genes & Development*, 12(18), pp.2852–2862. Available at: http://genesdev.cshlp.org/content/12/18/2852.full [Accessed June 20, 2015].

Porteus, M.H. & Baltimore, D., 2003. Chimeric Nucleases Stimulate Gene Targeting in Human Cells. *Science*, 300(5620), p.763. Available at: http://www.sciencemag.org/content/300/5620/763.short.

Potts, M.B. et al., 2015. Mode of action and pharmacogenomic biomarkers for exceptional responders to didemnin B. *Nature chemical biology*, advance on. Available at: http://dx.doi.org/10.1038/nchembio.1797 [Accessed April 16,

2015].

Qasim, M. et al., 2011. Differential proteome analysis of human embryonic kidney cell line (HEK-293) following mycophenolic acid treatment. *Proteome Sci*, 9, p.57.

Quadros, R.M. et al., 2015. Insertion of sequences at the original provirus integration site of mouse ROSA26 locus using the CRISPR/Cas9 system. *FEBS open bio*, 5, pp.191–197. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4372609&tool=pmcentrez&rendertype=abstract [Accessed April 3, 2015].

Ran, F.A., Hsu, P.D., et al., 2013. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*, 154(6), pp.1380–9. Available at: http://www.cell.com/article/S0092867413010155/fulltext [Accessed July 10, 2014].

Ran, F.A., Hsu, P.D., et al., 2013. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*, 8(11), pp.2281–2308.

Ran, F.A. et al., 2015. In vivo genome editing using Staphylococcus aureus Cas9. *Nature*, 520(7546), pp.186–191. Available at: http://dx.doi.org/10.1038/nature14299 [Accessed April 1, 2015].

Recchia, A. et al., 2004. Site-specific integration of functional transgenes into the human genome by adeno/AAV hybrid vectors. *Molecular therapy : the journal of the American Society of Gene Therapy*, 10(4), pp.660–70. Available at: http://dx.doi.org/10.1016/j.ymthe.2004.07.003 [Accessed February 8, 2016].

Recillas-Targa, F., 2004. Gene transfer and expression in mammalian cell lines and transgenic animals. *Methods in molecular biology (Clifton, N.J.)*, 267, pp.417–33. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15269440 [Accessed June 15, 2015].

van Rensburg, R. et al., 2013. Chromatin structure of two genomic sites for targeted transgene integration in induced pluripotent stem cells and hematopoietic stem cells. *Gene therapy*, 20(2), pp.201–14. Available at:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3661409&tool=pmcentrez&rendertype=abstract [Accessed August 26, 2015].

Reyon, D. et al., 2012. FLASH assembly of TALENs for high-throughput genome editing. *Nature biotechnology*, 30(5), pp.460–5. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3558947&tool=pmcentrez&rendertype=abstract [Accessed April 27, 2015].

Richard, G.-F., 2015. Shortening trinucleotide repeats using highly specific endonucleases: a possible approach to gene therapy? *Trends in genetics : TIG*. Available at: http://www.sciencedirect.com/science/article/pii/S0168952515000207 [Accessed March 9, 2015].

Rouet, P., Smih, F. & Jasin, M., 1994a. Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proceedings of the National Academy of Sciences*, 91(13), pp.6064–6068. Available at: http://www.pnas.org/content/91/13/6064.abstract.

Rouet, P., Smih, F. & Jasin, M., 1994b. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol. Cell. Biol.*, 14(12), pp.8096–8106. Available at: http://mcb.asm.org/content/14/12/8096.abstract?ijkey=ac69a4fc0e1844b02681b86de07ab83f4d7dfe1d&keytype2=tf_ipsecsha [Accessed June 8, 2015].

Rubin, G.M. et al., 2000. Comparative Genomics of the Eukaryotes. *Science*, 287(5461), pp.2204–2215. Available at: http://www.sciencemag.org/content/287/5461/2204.abstract.

Ruest, L.-B., Marcotte, R. & Wang, E., 2002. Peptide elongation factor eEF1A-2/S1 expression in cultured differentiated myotubes and its protective effect against caspase-3-mediated apoptosis. *The Journal of biological chemistry*, 277(7), pp.5418–25. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2803684&tool=pmcentrez&rendertype=abstract [Accessed August 2, 2015].

Sadelain, M., Papapetrou, E.P. & Bushman, F.D., 2012. Safe harbours for the integration of new DNA in the human genome. *Nature Reviews Cancer*, 12(1), pp.51–58. Available at: http://dx.doi.org/10.1038/nrc3179.

Saksouk, N., Simboeck, E. & Déjardin, J., 2015. Constitutive heterochromatin formation and transcription in mammals. *Epigenetics & chromatin*, 8, p.3. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4363358&tool=pmcentrez&rendertype=abstract [Accessed June 18, 2015].

Sakurai, K. et al., 2010. Efficient integration of transgenes into a defined locus in human embryonic stem cells. *Nucleic acids research*, 38(7), p.e96. Available at: http://nar.oxfordjournals.org/content/38/7/e96.long [Accessed June 8, 2015].

Sanber, K.S. et al., 2015. Construction of stable packaging cell lines for clinical lentiviral vector production. *Scientific reports*, 5, p.9021. Available at: http://www.nature.com/srep/2015/150312/srep09021/full/srep09021.html [Accessed July 23, 2015].

Sangiolo, D. et al., 2007. Lentiviral vector conferring resistance to mycophenolate mofetil and sensitivity to ganciclovir for in vivo T-cell selection. *Gene Ther*, 14(21), pp.1549–1554.

Sanjana, N.E. et al., 2012. A transcription activator-like effector toolbox for genome engineering. *Nature Protocols*, 7(1), pp.171–192. Available at: http://dx.doi.org/10.1038/nprot.2011.431.

Schiffer, J.T. et al., 2012. Targeted DNA Mutagenesis for the Cure of Chronic Viral Infections. *Journal of Virology*, 86(17), pp.8920–8936. Available at: http://jvi.asm.org/content/86/17/8920.abstract.

Schroeder, M.A. et al., 2006. Forced Expression of the “IY” Mutant Inosine Monophosphate Dehydrogenase II Results in Physiologically Significant Resistance to Mycophenolic Acid In Vitro. *ASH Annual Meeting Abstracts*, 108(11), p.5480. Available at: http://abstracts.hematologylibrary.org/cgi/content/abstract/108/11/5480

[Accessed May 19, 2015].

Schwank, G. et al., 2013. Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. *Cell stem cell*, 13(6), pp.653–8. Available at: http://www.cell.com/article/S1934590913004931/fulltext [Accessed July 12, 2014].

Shalem, O. et al., 2014a. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science (New York, N.Y.)*, 343(6166), pp.84–7. Available at: http://www.sciencemag.org/content/343/6166/84.short [Accessed July 10, 2014].

Shalem, O. et al., 2014b. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science (New York, N.Y.)*, 343(6166), pp.84–7. Available at: http://www.sciencemag.org/content/343/6166/84.full [Accessed July 10, 2014].

Shamovsky, I. et al., 2006. RNA-mediated response to heat shock in mammalian cells. *Nature*, 440(7083), pp.556–60. Available at: http://dx.doi.org/10.1038/nature04518 [Accessed May 18, 2015].

Shen, B. et al., 2014. Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat Methods*.

Shepherd, J.C. et al., 1989. Fruit flies with additional expression of the elongation factor EF-1 alpha live longer. *Proceedings of the National Academy of Sciences*, 86(19), pp.7520–7521. Available at: http://www.pnas.org/content/86/19/7520.abstract.

Shevchuk, N.A. et al., 2004. Construction of long DNA molecules using long PCR-based fusion of several fragments simultaneously. *Nucleic Acids Research*, 32(2), p.e19. Available at: http://nar.oxfordjournals.org/content/32/2/e19.abstract.

Shiina, N. et al., 1994. Microtubule severing by elongation factor 1 alpha. *Science (New York, N.Y.)*, 266(5183), pp.282–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7939665 [Accessed August 2, 2015].

Silva, G. et al., 2011. Meganucleases and other tools for targeted genome engineering: perspectives and challenges for gene therapy. *Current Gene Therapy*, 11(1), pp.11–27.

Sintchak, M.D. et al., 1996. Structure and Mechanism of Inosine Monophosphate Dehydrogenase in Complex with the Immunosuppressant Mycophenolic Acid. *Cell*, 85(6), pp.921–930. Available at: http://www.sciencedirect.com/science/article/pii/S0092867400812751.

Smith, C. et al., 2014. Whole-Genome Sequencing Analysis Reveals High Specificity of CRISPR/Cas9 and TALEN-Based Genome Editing in Human iPSCs. *Cell Stem Cell*, 15(1), pp.12–13. Available at: http://www.sciencedirect.com/science/article/pii/S1934590914002598.

Smith, J.R. et al., 2008. Robust, Persistent Transgene Expression in Human Embryonic Stem Cells Is Achieved with AAVS1-Targeted Integration. *STEM CELLS*, 26(2), pp.496–504. Available at: http://dx.doi.org/10.1634/stemcells.2007-0039.

Soares, D.C. et al., 2009. Structural Models of Human eEF1A1 and eEF1A2 Reveal Two Distinct Surface Clusters of Sequence Variation and Potential Differences in Phosphorylation. *PLoS ONE*, 4(7), p.e6315. Available at: http://dx.doi.org/10.1371/journal.pone.0006315.

Soares, D.C. & Abbott, C.M., 2013. Highly homologous eEF1A1 and eEF1A2 exhibit differential post-translational modification with significant enrichment around localised sites of sequence variation. *Biology direct*, 8(1), p.29. Available at: http://www.biologydirect.com/content/8/1/29 [Accessed July 30, 2015].

Söllü, C. et al., 2010. Autonomous zinc-finger nuclease pairs for targeted chromosomal deletion. *Nucleic acids research*, 38(22), pp.8269–76. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3001086&tool=pmcentrez&rendertype=abstract [Accessed August 7, 2015].

Spencer, S. et al., 2015. Stability of single copy transgene expression in CHOK1 cells is affected by histone modifications but not by DNA methylation. *Journal of biotechnology*, 195, pp.15–29. Available at: http://www.sciencedirect.com/science/article/pii/S0168165614010451 [Accessed February 23, 2015].

Sternberg, S.H. et al., 2014. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, 507(7490), pp.62–7. Available at: http://dx.doi.org/10.1038/nature13011 [Accessed July 9, 2014].

Strahl, B.D. & Allis, C.D., 2000. The language of covalent histone modifications. *Nature*, 403(6765), pp.41–5. Available at: http://dx.doi.org/10.1038/47412 [Accessed September 30, 2014].

Strathdee, D., Ibbotson, H. & Grant, S.G.N., 2006. Expression of transgenes targeted to the Gt(ROSA)26Sor locus is orientation dependent. *PloS one*, 1(1), p.e4. Available at: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0000004 [Accessed August 15, 2015].

Streubel, J. et al., 2012. TAL effector RVD specificities and efficiencies. *Nature Biotechnology*, 30(7), pp.593–595. Available at: http://dx.doi.org/10.1038/nbt.2304.

Strong, C.L. et al., 2015. Damaging the Integrated HIV Proviral DNA with TALENs. *PloS one*, 10(5), p.e0125652. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4422436&tool=pmcentrez&rendertype=abstract [Accessed May 18, 2015].

Sun, N. et al., 2012. Optimized TAL effector nucleases (TALENs) for use in treatment of sickle cell disease. *Molecular BioSystems*, 8(4), pp.1255–1263. Available at: http://dx.doi.org/10.1039/C2MB05461B.

Suster, M.L., Sumiyama, K. & Kawakami, K., 2009. Transposon-mediated BAC transgenesis in zebrafish and mice. *BMC genomics*, 10(1), p.477. Available at: http://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-10-477

[Accessed January 19, 2016].

Suzuki, K. et al., 2014. Targeted gene correction minimally impacts whole-genome mutational load in human-disease-specific induced pluripotent stem cell clones. *Cell stem cell*, 15(1), pp.31–6. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4144407&tool=pmcentrez&rendertype=abstract [Accessed August 7, 2015].

Suzuki, M.M. & Bird, A., 2008. DNA methylation landscapes: provocative insights from epigenomics. *Nature reviews. Genetics*, 9(6), pp.465–76. Available at: http://dx.doi.org/10.1038/nrg2341 [Accessed July 9, 2014].

Szczepek, M. et al., 2007. Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nature biotechnology*, 25(7), pp.786–93. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17603476 [Accessed July 26, 2015].

Tatsuka, M. et al., 1992. Elongation factor-1 alpha gene determines susceptibility to transformation. *Nature*, 359(6393), pp.333–336.

Tebas, P. et al., 2013. Antiviral effects of autologous CD4 T cells genetically modified with a conditionally replicating lentiviral vector expressing long antisense to HIV. *Blood*, 121(9), pp.1524–33. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3587318&tool=pmcentrez&rendertype=abstract [Accessed July 24, 2015].

Tebas, P. et al., 2014. Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. *The New England journal of medicine*, 370(10), pp.901–10. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4084652&tool=pmcentrez&rendertype=abstract [Accessed July 10, 2014].

Teschendorf, C. et al., 2002. Comparison of the EF-1 alpha and the CMV promoter for engineering stable tumor cell lines using recombinant adeno-associated virus. *Anticancer research*, 22(6A), pp.3325–30. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12530082 [Accessed June 21, 2015].

Throm, R.E. et al., 2009a. Efficient construction of producer cell lines for a SIN lentiviral vector for SCID-X1 gene therapy by concatemeric array transfection. *Blood*, 113(21), pp.5104–5110. Available at: http://bloodjournal.hematologylibrary.org/content/113/21/5104.abstract.

Throm, R.E. et al., 2009b. Efficient construction of producer cell lines for a SIN lentiviral vector for SCID-X1 gene therapy by concatemeric array transfection. *Blood*, 113(21), pp.5104–10. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2686181&tool=pmcentrez&rendertype=abstract [Accessed July 24, 2015].

Tiyaboonchai, A. et al., 2014. Utilization of the AAVS1 safe harbor locus for hematopoietic specific transgene expression and gene knockdown in human ES cells. *Stem cell research*, 12(3), pp.630–7. Available at: http://www.sciencedirect.com/science/article/pii/S1873506114000130 [Accessed February 7, 2016].

Tsai, S.Q. et al., 2014. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nature biotechnology*, 32(6), pp.569–76. Available at: http://dx.doi.org/10.1038/nbt.2908 [Accessed July 9, 2014].

Tuttle, J. & Krenitsky, T., 1984. Effects of acyclovir and its metabolites on purine nucleoside phosphorylase. *J. Biol. Chem.*, 259(7), pp.4065–4069. Available at: http://www.jbc.org/content/259/7/4065.abstract?sid=bc196a3b-11ce-42c4-9b4d-80f9b67c7156 [Accessed July 28, 2015].

Tuttle, J. V, Krenitsky, T.A. & Elion, G.B., 1983. Effects of acyclovir and its metabolites on hypoxanthine-guanine phosphoribosyltransferase. *Biochem Pharmacol*, 32(20), pp.3011–3015.

Uetsuki, T. et al., 1989. Isolation and characterization of the human chromosomal gene for polypeptide chain elongation factor-1 alpha. *The Journal of biological chemistry*, 264(10), pp.5791–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2564392 [Accessed August 1, 2015].

Urnov, F.D. et al., 2005. Highly efficient endogenous human gene correction using

designed zinc-finger nucleases. *Nature*, 435(7042), pp.646–651. Available at: http://dx.doi.org/10.1038/nature03556.

Valton, J. et al., 2012. Overcoming transcription activator-like effector (TALE) DNA binding domain sensitivity to cytosine methylation. *The Journal of biological chemistry*, 287(46), pp.38427–32. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3493886&tool=pmcentrez&rendertype=abstract [Accessed August 6, 2015].

Veeramah, K.R. et al., 2013. Exome sequencing reveals new causal mutations in children with epileptic encephalopathies. *Epilepsia*, 54(7), pp.1270–81. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3700577&tool=pmcentrez&rendertype=abstract [Accessed August 2, 2015].

Vera, M. et al., 2014. The translation elongation factor eEF1A1 couples transcription to translation during heat shock response. *eLife*, 3, p.e03164. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4164936&tool=pmcentrez&rendertype=abstract [Accessed August 2, 2015].

Veres, A. et al., 2014. Low Incidence of Off-Target Mutations in Individual CRISPR-Cas9 and TALEN Targeted Human Stem Cell Clones Detected by Whole-Genome Sequencing. *Cell Stem Cell*, 15(1), pp.27–30. Available at: http://www.sciencedirect.com/science/article/pii/S1934590914001866.

Viville, S., 1997. Transgenic Animals. In L. M. Houdebine, ed. *Transgenic Animal Generation and Use*. CRC Press, pp. 307–323. Available at: https://books.google.com/books?hl=en&lr=&id=E3H-XelIngkC&pgis=1 [Accessed June 23, 2015].

Wang, J. et al., 2012. Targeted gene addition to a predetermined site in the human genome using a ZFN-based nicking enzyme. *Genome Research*, 22(7), pp.1316–1326. Available at: http://genome.cshlp.org/content/22/7/1316.abstract.

Wang, X. et al., 2015. Unbiased detection of off-target cleavage by CRISPR-Cas9

and TALENs using integrase-defective lentiviral vectors. *Nature Biotechnology*, advance on. Available at: http://dx.doi.org/10.1038/nbt.3127 [Accessed January 19, 2015].

Weber, E. et al., 2011. A Modular Cloning System for Standardized Assembly of Multigene Constructs. *PLoS ONE*, 6(2), p.e16765. Available at: http://dx.doi.org/10.1371/journal.pone.0016765.

Wettstein, R., Bodak, M. & Ciaudo, C., 2015. Generation of a Knockout Mouse Embryonic Stem Cell Line Using a Paired CRISPR/Cas9 Genome Engineering Tool. *Methods in molecular biology (Clifton, N.J.)*. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25762293 [Accessed August 3, 2015].

Williams, A. et al., 2008. Position effect variegation and imprinting of transgenes in lymphocytes. *Nucleic acids research*, 36(7), pp.2320–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2367730&tool=pmcentrez&rendertype=abstract [Accessed June 21, 2015].

Williams, S. et al., 2005. CpG-island fragments from the HNRPA2B1/CBX3 genomic locus reduce silencing and enhance transgene expression from the hCMV promoter/enhancer in mammalian cells. *BMC biotechnology*, 5(1), p.17. Available at: http://bmcbiotechnol.biomedcentral.com/articles/10.1186/1472-6750-5-17 [Accessed February 7, 2016].

Wright, D.A. et al., 2014. TALEN-mediated genome editing: prospects and perspectives. *The Biochemical journal*, 462(1), pp.15–24. Available at: http://www.biochemj.org/content/462/1/15.abstract [Accessed April 2, 2015].

Wu, X. et al., 2006. Mycophenolic acid is a potent inhibitor of angiogenesis. *Arteriosclerosis, thrombosis, and vascular biology*, 26(10), pp.2414–6. Available at: http://atvb.ahajournals.org/content/26/10/2414.full [Accessed May 26, 2015].

Wu, Y. et al., 2013. Correction of a genetic disease in mouse via use of CRISPR-Cas9. *Cell stem cell*, 13(6), pp.659–62. Available at: http://www.cell.com/article/S1934590913004621/fulltext [Accessed July 16,

2014].

Wurm, F.M., 2004. Production of recombinant protein therapeutics in cultivated mammalian cells. *Nature biotechnology*, 22(11), pp.1393–8. Available at: http://dx.doi.org/10.1038/nbt1026 [Accessed July 13, 2014].

Xiao, A. et al., 2013. Chromosomal deletions and inversions mediated by TALENs and CRISPR/Cas in zebrafish. *Nucleic Acids Research*. Available at: http://nar.oxfordjournals.org/content/early/2013/06/06/nar.gkt464.abstract.

Xu, W., Russ, J.L. & Eiden, M. V, 2012. Evaluation of residual promoter activity in γ-retroviral self-inactivating (SIN) vectors. *Molecular therapy : the journal of the American Society of Gene Therapy*, 20(1), pp.84–90. Available at: http://dx.doi.org/10.1038/mt.2011.204 [Accessed August 10, 2015].

Yan, B., Li, D. & Gou, K., 2010. Homologous illegitimate random integration of foreign DNA into the X chromosome of a transgenic mouse line. *BMC molecular biology*, 11(1), p.58. Available at: http://www.biomedcentral.com/1471-2199/11/58 [Accessed June 22, 2015].

Yan, B.-W. et al., 2013. Mechanism of random integration of foreign DNA in transgenic mice. *Transgenic research*, 22(5), pp.983–92. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23483296 [Accessed June 22, 2015].

Yang, H. et al., 2013. One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell*, 154(6), pp.1370–9. Available at: http://www.sciencedirect.com/science/article/pii/S0092867413010167 [Accessed July 9, 2014].

Yang, M. et al., 2014. CRISPR/Cas9 mediated generation of stable chondrocyte cell lines with targeted gene knockouts; analysis of an aggrecan knockout cell line. *Bone*, 69, pp.118–25. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25260929 [Accessed August 3, 2015].

Yi, Y., Noh, M.J. & Lee, K.H., 2011. Current advances in retroviral gene therapy. *Current gene therapy*, 11(3), pp.218–28. Available at:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3182074&tool=pmcentrez&rendertype=abstract [Accessed February 8, 2016].

Yu, C. et al., 2015. Small Molecules Enhance CRISPR Genome Editing in Pluripotent Stem Cells. *Cell Stem Cell*, 16(2), pp.142–147. Available at: http://www.sciencedirect.com/science/article/pii/S1934590915000041 [Accessed February 6, 2015].

Yusa, K. et al., 2011. Targeted gene correction of α1-antitrypsin deficiency in induced pluripotent stem cells. *Nature*, 478(7369), pp.391–4. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3198846&tool=pmcentrez&rendertype=abstract [Accessed July 28, 2015].

Zambrowicz, B.P. et al., 1997. Disruption of overlapping transcripts in the ROSA βgeo 26 gene trap strain leads to widespread expression of β-galactosidase in mouse embryos and hematopoietic cells. *Proceedings of the National Academy of Sciences*, 94(8), pp.3789–3794. Available at: http://www.pnas.org/content/94/8/3789.abstract.

Zhang, F. et al., 2010. A ubiquitous chromatin opening element (UCOE) confers resistance to DNA methylation-mediated silencing of lentiviral vectors. *Molecular therapy : the journal of the American Society of Gene Therapy*, 18(9), pp.1640–9. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2956914&tool=pmcentrez&rendertype=abstract [Accessed December 8, 2015].

Zhang, J.-H. et al., 2014. Improving the specificity and efficacy of CRISPR/CAS9 and gRNA through target specific DNA reporter. *Journal of biotechnology*. Available at: http://www.sciencedirect.com/science/article/pii/S0168165614008384 [Accessed September 8, 2014].

Zhang, S. et al., 2014. TALEN mediated somatic mutagenesis in murine models of cancer. *Cancer Research*. Available at: http://cancerres.aacrjournals.org/content/early/2014/07/25/0008-5472.CAN-14-0529.abstract.

Zheng, H. & Wilson, J.H., 1990. Gene targeting in normal and amplified cell lines. *Nature*, 344(6262), pp.170–3. Available at: http://dx.doi.org/10.1038/344170a0 [Accessed July 28, 2015].

Zhou, J. et al., 2014. Dual sgRNAs facilitate CRISPR/Cas9-mediated mouse genome targeting. *FEBS Journal*, 281(7), pp.1717–1725. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24494965 [Accessed July 23, 2015].

Zhou, X. et al., 2015. Generation of CRISPR/Cas9-mediated gene-targeted pigs via somatic cell nuclear transfer. *Cellular and molecular life sciences : CMLS*, 72(6), pp.1175–84. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25274063 [Accessed August 3, 2015].

Zhou, Y. et al., 2014. High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature*, 509(7501), pp.487–91. Available at: http://dx.doi.org/10.1038/nature13166 [Accessed July 13, 2014].

Zincarelli, C. et al., 2008. Analysis of AAV serotypes 1-9 mediated gene expression and tropism in mice after systemic injection. *Molecular therapy : the journal of the American Society of Gene Therapy*, 16(6), pp.1073–80. Available at: http://dx.doi.org/10.1038/mt.2008.76 [Accessed September 18, 2015].

Zou, J. et al., 2011. Oxidase-deficient neutrophils from X-linked chronic granulomatous disease iPS cells: functional correction by zinc finger nuclease–mediated safe harbor targeting. *Blood*, 117(21), pp.5561–5572. Available at: http://bloodjournal.hematologylibrary.org/content/117/21/5561.abstract.